

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 08:50:49 / Search time 60 Seconds
(without alignments)
2670.074 Million cell updates/sec

Title: US-10-656-499-2

Perfect score: 2993

Sequence: 1 MAGRLTWISFIVGALDSD.....QFDMVPLVILRLRSVTCDD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2993	100.0	567	5	AAE24846 KSHV late
2	125	4.2	777	4	ABE60323 Drosophil
3	122	4.1	364	3	AAB53186 Macaca mu
4	121.5	4.1	2518	3	AAB40574 Human ORF
5	120	4.0	253	3	AAB53182 Macaca mu
6	117	3.9	1316	4	ABG22997 Novel hum
7	115	3.8	1007	4	AAW78838 Human pro
8	115	3.8	2507	6	ABU61812 Human nuc
9	114.5	3.8	1321	4	ABG25416 Novel hum
10	113.5	3.8	1005	7	AD43360 Chlamydia
11	113	3.8	1007	6	ABP96241 Human pro
12	113	3.8	1007	7	ADC06835 Human pro
13	113	3.8	1019	4	AAW79822 Human pro
14	112.5	3.8	466	7	Ades5522 Rat Prote
15	112.5	3.8	466	7	Ades5524 Human pro
16	112	3.7	1007	7	ADC06800 Human pro
17	111	3.7	509	6	ABU34009 Protein e
18	110.5	3.7	434	3	AAAG50910 Arabidops
19	110	3.7	975	4	ABE63940 Drosophil
20	108.5	3.6	503	7	Ades58665 Human pro
21	108.5	3.6	503	7	Ades58669 Human pro
22	108.5	3.6	550	4	AAW94880 Human pro
23	108.5	3.6	561	4	AAW40771 Human pol
24	108	3.6	953	6	ABU36789 Protein e
25	107.5	3.6	407	3	AAAG50911 Arabidops

ALIGNMENTS

RESULT 1

AAE24846

ID AAE24846 standard; protein; 567 AA.

XX AC AAE24846;

XX AC

DT 30-OCT-2002 (first entry)

XX XX

DE KSHV latency-associated nuclear antigen 2 (LANA2).

XX XX

KW Kaposi's sarcoma-associated herpesvirus; KSHV; Castlemann's disease; latency-associated nuclear antigen 2; LANA2; apoptosis; tonsil tissue; primary effusion lymphoma; lymph node; spleen; skin lesion; virucide; cytostatic; antibody-based therapy; antisense therapy.

XX OS Human herpesvirus 8.

XX FN WC200245744-A1.

XX PD 13-JUN-2002.

XX PF 07-DEC-2001; 2001WO-US047217.

XX PR 08-DEC-2000; 2000US-00733728.

XX PA (UYCO) UNIV COLUMBIA NEW YORK.

XX PI Chang Y, Moore PS;

XX DR WPI; 2002-519642/55.

XX DR N-PSDB; AAD40494.

XX PT Novel nucleic acid encoding Kaposi's sarcoma-associated herpesvirus

XX PT latency-associated nuclear antigen 2 polypeptide for inhibiting p53

XX PS mediated apoptosis of a cell.

XX PS Claim 47; Fig 13; 11pp; English.

XX CC The invention relates to an isolated nucleic acid which encodes a

XX CC Kaposi's sarcoma-associated herpesvirus (KSHV) latency-associated nuclear

XX CC antigen 2 (LANA2) polypeptide or its fragment. The invention is useful

XX CC for inhibiting p53 mediated apoptosis of a cell, for immortalising a

XX CC cell, preferably haematopoietic tissue cell or B cell and for producing

XX CC an antibody. The invention is useful for treating Kaposi's sarcoma-

XX CC associated herpesvirus infection in a subject. The invention is also is

XX CC useful for determining whether a subject, including mouse, rat, dog,

XX CC guinea pig, ferret, rabbit, primate, or human being is afflicted with a

XX CC disease associated with Kaposi's sarcoma-associated herpesvirus infection

26 107.5 3.6 433 3 AAG50896 Arabidops
27 107 3.6 417 4 AAE13003
28 107 3.6 883 4 AAB61242 Murine br
29 107 3.6 883 6 ABO32678 Secreted
30 107 3.6 883 7 ADB90787 Mouse bre
31 107 3.6 883 7 AAR98747 P. vivax
32 107 3.6 1018 2 AAW97039 A. secreta
33 107 3.6 1018 4 AAG66528 Plasmodiu
34 107 3.6 1033 5 ABP66122 Bifidobac
35 107 3.6 1061 4 ABB59318 Drosophil
36 106.5 3.6 686 5 AAE16790 Human tra
37 104.5 3.5 503 2 AAY15103 Modified
38 104.5 3.5 1973 7 ADC06828 Human pro
39 104.5 3.5 2047 5 AAU75883 Human adh
40 104.5 3.5 3863 7 ADC06827 Human pro
41 104 3.5 166 4 AAM39445 Human pol
42 104 3.5 170 4 AAG64532 Human zin
43 104 3.5 175 5 ABP41810 Human ova
44 104 3.5 195 4 AAU15901 Human nov
45 104 3.5 195 6 ABU54970 Human nov

CC of a B cell, such as Castleman's disease and primary effusion lymphoma or
 CC whether a subject is infected with KSHV, by contacting tonsil tissue,
 CC lymph nodes, spleen, skin lesions, blood, serum, plasma cerebrospinal
 CC fluid, lymphocytes, urine, transudates, exudates, bone marrow cells, and
 CC detecting any antibody which is bound to LAN22, by ELISA, IFA or Western
 CC blotting. The invention is also useful for treating B-cell associated
 CC diseases such as Castleman's disease, and primary effusion lymphoma. The
 CC invention is useful in antibody-based therapy and antisense therapy. The
 CC present sequence is Kaposi's sarcoma-associated herpesvirus (KSHV)
 CC latency-associated nuclear antigen 2 (LANA2)
 XX
 SQ Sequence 567 AA;

Query Match 100.0%; Score 2993; DB 5; Length 567;
 Best Local Similarity 100.0%; Pred. No. 3.2e-295;
 Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAGRLTWISSEFVIGALDSKYPVWLDKSTGTFLAPARNDVPLDSLQFFIDFKKEC 60
 Db 1 MAGRLTWISSEFVIGALDSKYPVWLDKSTGTFLAPARNDVPLDSLQFFIDFKKEC 60
 QY 61 LSKGLHPRDLGSPITAFKICTTSRRLRLPGEEVVGGINCRWRLLCAEVKECWWC 120
 Db 61 LSKGLHPRDLGSPITAFKICTTSRRLRLPGEEVVGGINCRWRLLCAEVKECWWC 120
 QY 121 VHARTLHSGSSILWEILYOHVSLEKRRRPRRPFVGENSDSEEDHAFCDVPVTTQGA 180
 Db 121 VHARTLHSGSSILWEILYOHVSLEKRRRPRRPFVGENSDSEEDHAFCDVPVTTQGA 180
 QY 181 ESDSGDEGPFSTRHSASGVQVDDANADSPGSDGPFSTRHSASGVQVDDANADSPGSDG 240
 Db 181 ESDSGDEGPFSTRHSASGVQVDDANADSPGSDGPFSTRHSASGVQVDDANADSPGSDG 240
 QY 241 DDLTLDKESACALMTHVGOEMDMLRAMCDEDLFDLIGIPEDVIATSQPGDITDASGV 300
 Db 241 DDLTLDKESACALMTHVGOEMDMLRAMCDEDLFDLIGIPEDVIATSQPGDITDASGV 300
 QY 301 TEGSIAASAVGAGVEDVYLAGALEAQNAGVEYVLEISDEVEDDAGLPPASRRRPVWGEF 360
 Db 301 TEGSIAASAVGAGVEDVYLAGALEAQNAGVEYVLEISDEVEDDAGLPPASRRRPVWGEF 360
 QY 361 LWDGGRPRRPTTRIRKRLRSAYRVARPPVMTDLGLVEYVYFGRPAMSLVERKV 420
 Db 361 LWDGGRPRRPTTRIRKRLRSAYRVARPPVMTDLGLVEYVYFGRPAMSLVERKV 420
 QY 421 FILCSQNLADISHSLHRSKGLRVLLPKPDNNTPGPDVNLAAVLRSFASGLVIVSLR 480
 Db 421 FILCSQNLADISHSLHRSKGLRVLLPKPDNNTPGPDVNLAAVLRSFASGLVIVSLR 480
 QY 481 SGIVYKQLCKSTVLHGNNPPKPGVICGLSSRAVLDFVFNVAQYRIQGHIEHKTTVFIG 540
 Db 481 SGIVYKQLCKSTVLHGNNPPKPGVICGLSSRAVLDFVFNVAQYRIQGHIEHKTTVFIG 540
 QY 541 GPDTSAEQFDMVPLVILKRLRSVTCDD 567
 Db 541 GPDTSAEQFDMVPLVILKRLRSVTCDD 567

RESULT 2

ABB60323
 ID ABB60323 standard; protein; 777 AA.

XX AC ABB60323;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 7761.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.
 XX OS Drosophila melanogaster.

PN WO200171042-A2.
 XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US0009231.
 PF 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX N-ESDB; ABL04426.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID NO 7761; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 777 AA;

Query Match 4.2%; Score 125; DB 4; Length 777;
 Best Local Similarity 23.5%; Pred. No. 0.0078;
 Matches 80; Conservative 35; Mismatches 94; Indels 132; Gaps 20;
 QY 93 GEYEVVVGINCRRWRLLCAEVK-----RCWCVHARTLHSGSSILWEILYQHSVR 143
 Db 102 GAELYKQEAAYR-RHCKTKCGEKLPREESRPMCKCY---TRFSSASN----- 146
 QY 144 LEKHRR-----RPRRPVVG-----ENSDSSEEDHPAFCDVPVTTQGA 181
 Db 147 LSKHRRSRPTCCQPEVDSPGSDGMKCHKAFKKDRNRDSDDED-----TTSEE 196
 QY 182 SEDSGDEGPFSTRHSASGVQVDDANADSPGSDGEGPSTRHSASQ-----PPPADE 231
 Db 197 SEDSDDDIPLASRLKTKLKQ-ESQNSD---SGDECFDFEPNNSDDADASGFLQPPPA-- 250
 QY 232 TTVHTDNVEDDLTLDKESACALMTHVGOEMDMLRAMCDEDLFDLL-----GI-PEDVI 285
 Db 251 -WVKVEAFDEE---DFEYQDASMYVKTSTDFISNE--KDKLLDVLNKGDLKPFESL 303
 QY 286 ATSQPGDITDASGVVTEGSIASAVAGVEDVYLAGALEAQNAGVEYVLEISDEVEDDGA 345
 Db 304 KVEQ-----GAGILDE--IAAVPLVEVAEDVLELRGHQMEK-- 338
 QY 346 GLPPASRRRPVVGELWDDGPREHERPTTTRI-----RHRK 381
 Db 339 --PP-----GPRKGRPPKEKIPVVRKTKRK 362

RESULT 3

AAB53186
 ID AAB53186 standard; protein; 364 AA.

XX AC AAB53186;

XX DT 28-FEB-2001 (first entry)

XX OS Macaca mulatta rhadinovirus 17577 RRV R13 protein SEQ ID NO:129.

XX	Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
XW	genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
XX	ii-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
XW	cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW	lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW	splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
XX	
OS	Macaca mulatta rhadinovirus 17577.
XX	
XX	W0200028040-A2.
PN	
XX	
PD	18-MAY-2000.
XX	
XX	
PF	05-NOV-1999; 99WO-US026260.
XX	
PR	06-NOV-1998; 98US-0107507P.
PR	20-NOV-1998; 98US-0109409P.
XX	
XX	(UYOR-) UNIV OREGON HEALTH SCI.
PA	
XX	
PI	Wong SW, Axthelm MK, Searles RP;
XX	
DR	WPI; 2000-376552/32.
XX	
XX	New rhesus rhadino virus for producing non-human primate model useful for
PT	testing potential treatments and efficacy of the candidate vaccine for
PT	conditions associated with RRV infection.
XX	
PS	Claim 5; Page 135; 141pp; English.
XX	
CC	The present invention describes a novel rhesus macaque rhadinovirus
CC	called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC	RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC	encoded by the genome sequence. The present invention also specifically
CC	claims the individual open reading frame (ORF) nucleotide sequences from
CC	the genome which encode the individual proteins, but these sequences are
CC	not given. A non-human animal infected with RRV can be used for testing
CC	the efficacy of drug in the treatment of condition associated with
CC	infection with RRV such as Kaposi's sarcoma, lymphoproliferative
CC	disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
CC	hypergammaglobulinemia or autoimmune haemolytic anaemia, by
CC	administering the drug to a immuno-compromised non-human primate
CC	preferably Rhesus macaque monkey obtained by as a result of infection by
CC	Simian Immunodeficiency Virus (SIV). RRV is useful for producing non-
CC	human primate model for testing potential treatments for conditions
CC	associated with RRV infection. It is also useful for testing the efficacy
CC	of the candidate vaccine against RRV infection or conditions associated
CC	with its infection by administering the vaccine to the subject capable of
CC	infection with RRV, inoculating the subject with RRV and observing the
CC	effect of vaccine. AAC64755 to AAC64765 and AAB53205 to AAB53213
CC	represent sequence used in the exemplification of the present invention
XX	
SQ	Sequence 364 AA;

```

Query Match      4.1%; Score 122; DB 3; Length 364;
Best Local Similarity 22.5%; Pred. NO. 0.0047;
Matches 82; Conservative 43; Mismatches 127; Indels 112; Gaps 19

QY      246 LDKESACALMYHVGOEMDLRMARCDLFLQLGIPEDVIATSQPGDITDASGVTGEIS 305
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      57 LDRDLECGRHQNVSECRKLLRLVRENAGFE---QDDARATTTFEG----- 99

QY      306 AASAVGAGVEDVYLAGALEACQNVAGEYVLETSDSEV-----DDGAGLPPASRR 353
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      100 -----GERFFYLPAVDPLCYA---CLDHSHTSVLNYLEAACVHGLEBFTPLPP--- 146

QY      354 RPVVGFLWDDGFRHREPTTRIRHKRLRSAYYRVAR-----PPVMITD---RLGVGV 404
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      147 -PAPAE---ADG-----AARSVYARAARLATVAPPHFDQITTFPWLRLRV 187

QY      405 FYFGRPAMSLEYE-----RKVFILCONPLADISHSLRSKGLRVLLPKPD-ONNTGP 457
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db	198	PFYG-----SLVAEHTSDRRGVRHLHKRDQKP--GHECFYG-TAYKMWLPKQLDGLPTP	244
Qy	458	GDVLLAAVLSFASGLVIVLSRGIYVKNLCKSTVLVHGNNP-----PKKF	504
Db	241	EQRETVCEIINGCEGVFLHGNELGMYDNRTRHTVRCAGNDAEGNHAQRAVRSSVKSOI	300
Qy	505	GVIICLSRAVLDPVFNVAQYRIQGH-HIKTTVTFIGDQPSAEGDMVPLVIKURLRSV	563
Db	301	FYVNGLLR-----LARSVPDGTVPESNAVTLYLGRPGSSKR-PQVPVTL-----V	346
Qy	564	TCDD	567
Db	347	ICQD	350
RESULT 4			
AAB40574			
ID	AAB40574 standard; protein; 2518 AA.		
XX	AC	AAB40574;	
XX	DT	08-FEB-2001 (first entry)	
XX	DE	Human ORFX ORF338 polypeptide sequence SEQ ID NO:676.	
XX	KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;	
XX	KW	vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	
XX	KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;	
XX	KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;	
XX	KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;	
XX	KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;	
XX	KW	antinaemic; gene therapy; cancer; proliferative disorder; hypertension;	
XX	KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	
XX	KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	
XX	KW	cholesterol ester storage; systemic lupus erythematosus; infection;	
XX	KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	
XX	KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	
XX	KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;	
XX	KW	thrombosis; contraceptive.	
XX	OS	Homo sapiens.	
XX	FN	WO2000058473-A2.	
XX	PD	05-OCT-2000.	
XX	PF	31-MAR-2000; 2000WO-US008621.	
XX	PR	31-MAR-1999; 99US-0127607P.	
XX	PR	02-APR-1999; 99US-0127636P.	
XX	PR	05-APR-1999; 99US-0127728P.	
XX	PR	30-MAR-2000; 2000US-00540763.	
XX	PA	(CURA-) CURAGEN CORP.	
XX	FI	Shimkets RA, Leach M;	
XX	DR	WPI; 2000-602362/57.	
XX	DR	NI-PSDE; AAC74783.	
XX	PT	Novel nucleic acids and peptides derived from open reading frame X,	
XX	PT	useful for treating e.g. cancers, proliferative disorders,	
XX	XX	neurodegenerative disorders and cardiovascular disease.	
XX	PS	Claim 11; Page 784-790; 5507pp; English.	
XX	CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,	
XX	CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX	
XX	CC	sequences have activities such as: cytostatic; hepatotropic; vulnery;	
XX	CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;	
XX	CC	anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;	
XX	CC	cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;	
XX	CC	dermatological; immunosuppressive; antiinflammatory; antibacterial;	

XX OS Homo sapiens.
 XX WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US008631.
 XX PR 31-MAR-2000; 2000US-00540217.
 XX PR 23-AUG-2000; 2000US-00649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Dmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS87184.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX PS Claim 20; SEQ ID NO 53356; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 1316 AA;
 Query Match 3.9%; Score 117; DB 4; Length 1316;
 Best Local Similarity 19.7%; Pred. No. 0.12;
 Matches 122; Conservative 55; Mismatches 187; Indels 256; Gaps 29;
 QY 24 LVKWLDRSTGTFAPARNVDVFLDSLOFIDFK-----RECL-----SKGLHPRDLG 72
 DB 276 LYAW-DNRRETFRKSLGNDETDEKXKFLGFFKVKRNSKGCCLTTPNSPMHRSRLT 334
 QY 73 SPITAFGKICTTS-----RLRLRLPGEYEVVQGINCRWELLCAEYKECWCWCHART 125
 DB 335 GPSLSLGSISGVSKEMKRRAPPPGSGPPVQD-----KASE 373
 QY 126 HLHSGSLWEILLQHSVRLKRR-----PRRFVGENSSDSEHHPA-FCQVP 174
 DB 374 KVSLSGS-----QIDLOKKRRAPAPPPPPPPSPILPNRTEDKEENRKRSTVSLP 424
 QY 175 VTGTGAESEDSGEGSTRHSASGVQVDDANADSPGSDGEGSTRHSQPPADETTV 234
 DB 425 L-----GSGSHCSPPGAPQVLSAEETVSGSCFASDTEEDSGVMSFSD-----IV 472
 QY 235 HTDNVEDDLTLLDK-----ESACALMYHV----- 259
 DB 473 SLDSQQDSMKYKDKWATQDCQDLQAGTPDLGPKQSPLEKNGSENSHLRTEKAVTAS 532

QY 260 -QEMDLAR-----AMCDEDLFLLGIPEDVIATSQ-----PG 291
 DB 533 NDEEDLLIAGEPRKTLAELDEDEEM-----EDSYETDSSLTSSIHGASNHCPQDAMIPH 588
 QY 292 GDTASGVVTEGSIASAVGAGVEDVYLAGALEQNAVAGEVYLEISDEVEDDGAGLPAS 351
 DB 589 GDTDAIPVTFFIGESDDPVDGSLFS-----NRNNAGSF-----DSEGV--AS 629
 QY 352 RR---RPVVGEFLMDGPRRHERPTRRIRHKLRKSAYRVARPPVMTDRLGVEVYFG 408
 DB 630 RDSLAPLQAE-----HSQP-----HEKARE-----EV----- 652
 QY 409 RPAMSEVERKVFILCSQNPDLADISHSLHS-RKGLRVLLFKPDDNNTGPGDVNLLAVL 467
 DB 653 -PAL-----HPASHDVGKIRVAL-----SNISKQGNLMETAPRV 686
 QY 468 RSPASGL-----VIVSLRSGIV-----VKNLCKSTVLYH-- 496
 DB 687 TSFASNLHTDNLNAKVKKVYGCADGERTQATERNVSNQVNEKDSNDKNAALAPTSHQR 746
 QY 497 GNNPPKFGVICGLSSRAVL 516
 DB 747 GQPGKSYRLKHGLTYKII 766
 RESULT 7
 AAM78838
 ID AAM78838 standard; protein; 1007 AA.
 XX AAM78838;
 XX 06-NOV-2001 (first entry)
 XX Human protein SEQ ID NO 1500.
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 XX WO200157190-A2.
 XX 09-AUG-2001.
 XX 05-FEB-2001; 2001WO-US004098.
 XX 03-FEB-2000; 2000US-00496914.
 XX 27-APR-2000; 2000US-00560875.
 XX 20-JUN-2000; 2000US-00598075.
 XX 19-JUL-2000; 2000US-00620325.
 XX 01-SEP-2000; 2000US-00654936.
 XX 15-SEP-2000; 2000US-00663561.
 XX 20-OCT-2000; 2000US-00693325.
 XX 30-NOV-2000; 2000US-00728422.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao Qa, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 XX N-PSDB; AAK51971.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX Claim 20; Page 3790-3792; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM79323-AM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulatory
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
XX Sequence 1007 AA:
SQ

[illegible]

RESULT 8	
ABU61812	
ID	ABU61812 standard; protein; 2507 AA.
XX	
AC	ABU61812;
XX	
DT	12-AUG-2003 (first entry)
XX	
DE	Human nuclear receptor corepressor SMRTe.
XX	
KW	Human; SMRTe; nuclear receptor corepressor; gene therapy; tissue typing;
KW	cancer.
XX	

OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FF	Domain
FT	166..429
FT	/label= SNC_domain
FT	Region
FT	166..206
FT	/label= Amphipathic helix
FT	/note= "Residues 172, 179, 186, 194 and 200 make up the hydrophobic heptad repeat"
FT	430..486
FT	Domain
FT	/label= SANT_A_domain
FT	613..669
FT	/label= SANT_B_domain
XX	
PN	US2003027137-A1.
XX	
PD	06-FEB-2003.
XX	
PF	27-MAR-2001; 2001US-00819104.
XX	
PR	29-MAR-2000; 2000US-0193138P.
XX	
PA	(CHEN/) CHEN J D.
XX	
PI	Chen JD;
XX	
PI	WPI: 2003-466139/44.
DR	N-PSTB; AC62249, ACA62250.
DR	
XX	
PT	New SMRte proteins and nucleic acids, useful in gene therapy, predictive medicine, therapeutic or prophylactic treatment, chromosome mapping, tissue typing and in forensic biology.
PT	
XX	
PS	Claim 9; Fig 1; 90pp; English.
XX	
CC	The invention relates to an isolated SMRte nucleic acid molecule. The nucleic acids are useful in gene therapy, as hybridisation probes for identifying SMRte-encoding nucleic acid molecules and as primers for amplifying of SMRte nucleic acid molecules. The polypeptides are useful as immunogens to raise anti-SMRte antibodies. The SMRte molecules are useful as targets for discovering and developing modulating agents to regulate a variety of cellular processes, in screening assays, in predictive medicine, in therapeutic or prophylactic treatment, in chromosome mapping, tissue typing and in forensic identification of a biological sample. Modulators of SMRte are useful for treating or preventing a condition associated with aberrant SMRte protein or nucleic acid expression or activity, such as cancer. The present sequence represents the amino acid sequence of the human nuclear receptor corepressor SMRte
XX	
SQ	Sequence 2507 AA;
	Query Match 3.8%; Score 115; DB 6; Length 2507;
	Best Local Similarity 21.0%; Pred. No. 0.52;
	Matches 75; Conservative 53; Mismatches 139; Indels 90; Gaps 16;
QY	105 RNRLLCAEV--KECWCVIARTHLHSGSLMEILLYQHVSVRLEKHRRRPFRVGENSDS 162
Db	631 RNWSAIARMVGSKTVSQCKNFYNFKKRNQLDEILLQQHKLVKEKNARKKKAPAAAS 690
QY	163 SEEDHPAFCDVPVTO-----TGAESDSDGPGSTFHSASGVQ---PVDDANADS 209
Db	691 EBAAFP-----PVEDENEASGVONE-EEMVEEAEATVNSSDTESIPSHTEAKOT 744
QY	210 PGSGDEGPSTRHDSQP-----PPADETTVHTDNV-EDDLTL----- 246
Db	745 QNGPKFPATLGADGPPGPPPTPPEDIPAPTETSPASEATLATPPAPFPFPSPPPVV 804
QY	247 -----DKESACALMHVVGEMDMRAMCEDLFDLLGLPEDVIATS-----QPGG 292
Db	805 PKEEKEETAAPPVEEGE-----QKPFAEEHLAVDTKAESPVKSECTEAEEGPAKG 860
QY	293 DTDCAGVTTEGSIASAAGVEDVYLACALEAQNAG-----EYYLEISDSDEVDGAG- 346

QY	SQ	Sequence 2507 AA;	Query Match	Best Local Similarity	3.8%; Score 115; DB 6; Length 2507;	21.0%; Pred. No. 0.52;	Mismatches 53; Conservative 139; Indels 90; Gaps 16;
QY	105	RWRLLCAV--KECWCVHARTLHSGSLWEILYQHSVRLEKRRRRPRRPVGENSDS	162				
Db	631	RWSAIAARMVGSKTVSQCNFVYKKRQNLDEILQQHKLKVEKERNARKKKKAPAAAS	690				
QY	163	SEEDHPAFCDVPVTO-----TGAESDSGDGSPSTHSASGVQ-----PVDDANADS	209				
Db	691	EEAAPP-----PVVEDEENASGVGTGNE--EEMVEEAEATVNNSSDPTESIPSPHTEAAKOT	744				
QY	210	PGSGDGGPSTRSDSQF-----PPADETTVHTDNV--EDDLTLL-----	246				
Db	745	QONGPKPPATLGADGPPGPPPTPPEDIPAPTESIPASEATLAPTTPPPAPPPSSPPPVV	804				
QY	247	-----DKESACALMTHVGEMDMLRAMCDLFDLLGIPEDVIATS-----QPGG	292				
Db	805	PXKEKEETAAAPPVEEGSE-----QKPPAAEELAVDVTGKAESPFVKSCTEABEGPAKOK	860				
QY	293	DTDASGVVTEGSIASAASVAGGVEDVYLAGALBAQNNAV-----EYVLEISDEVIDDGAG-	346				

DB 861 DAAEAAEATFAELKAEKKEG-----SGRATTAKSGAGPOQSDSSATSCADEVDEAEGG 914
 QY 347 -----LPPASRRRPVVGDFLWDDGPRRHERP-TTTRIRHRKRLRSAYVYRAPPVMIT 397
 DB 915 DKNLLSRPRLTFTGD-----PRNASQKPLDLKQKORAA-----AIPPIQVT 961

RESULT 9

ABG25416
 ID ABG25416 standard; protein; 1321 AA.

XX AC ABG25416;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #25407.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS89603.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX PS Claim 20; SEQ ID NO 55775; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1321 AA;

Query Match

Best Local Similarity 3.8%; Score 114.5; DB 4; Length 1321;

Matches 118; Conservative 61; Mismatches 213; Indels 213; Gaps 26;

QY 24 LVKWLDRSTGTFLAARNVDVPLDSLOFFIDFK-----RECL-----SKGLHPRDLIG 72
 DB 279 LYAW-DNRRETRKSLGNDKDETDKXKFLGFKVKNRNSKCLTTPNSPMHSRLTL 337
 QY 73 SPITAFKICTTYS-----RRLRLPGBEYFVVGINCRRWRLLCAEYKECWCVHART 125
 DB 338 GPSLSLGSISGVSVKSEMKRRAPPVPGSGPPVQD-----KASE 376
 QY 126 HLHSGSSLWEILYQHSVRLEKRR-----PRPFVGENSDSSSEEDHPA-PCDVP 174
 DB 377 KVSLS-----QIDLQKKRRAPAPPPPPPPPLIPNRTEDKEENRKSXTWVSLP 427
 QY 175 VTQTGAESDSDGEGFSTRHSASGVQVDDANADSPGSGDEGFSTRHSQPPPADETTV 234
 DB 428 L---GSGSHCSPDGAPOVLSAEETVSVGSCFASDITTEDSGVMSSPSD-----IV 475
 QY 235 HTDNVEDDLTLDK-----ESACALMVHVG-----259
 DB 476 SLDSQDSMKYDKWATDOEDCSDQLAGTDLGPQKPLWEKNGSENSHLTEKAVTAS 535
 QY 260 -QEMDMLAR-----AMCDEDLFDLLGIPEDVIATSQ-----PG 291
 DB 536 NDEEDLLIAGEFRKTLAELDEDELEEM-----EDSYETDTSLSITSGASNHCPQDAMIH 591
 QY 292 GDTDASGVVTGSGIAASAVGAVGVYLAGALEAONVAGEYVLEISDEVDGAGLPPAS 351
 DB 592 GDTDAIPVTFICEVSDDPVDSGLFS-----NRNNAGSF-----DSEGV--AS 632
 QY 352 RR---RPVVGEFLWDDGPRRHERPTTTRIRH---BKLSAYYRVARPPVMI-----396
 DB 633 RRDLSAPLOAHSQPHKAREEVPALHPASHDVGGIRVALSNISKDGNLMTAPRTHN 692
 QY 397 -----TDRLGVEV--FYFG-----RPAMSLVERKVFILC-----SQ 426
 DB 693 FALNLHTDNLNAKVKDKVYGCADGERTQATERVNSQPVNEKDSNDKNAALAPTSHQRQ 752
 QY 427 NFLADISHSLHSRKLRLVLLPKPD-----DNNTGPDVNLAAVLRSFASGLVIVS 478
 DB 753 NP--GKSLRKHLGLTYKLIIPPKSEMRCVDRDVSLSUGAIDELGNLVSHPATGIRIIS 810
 QY 479 LRSGI 483
 DB 811 LSSSV 815

RESULT 10

ADD43960
 ID ADD43960 standard; protein; 1005 AA.

XX AC ADD43960;

XX DT 15-JAN-2004 (first entry)

XX DE Chlamydia trachomatis immunogenic protein, SEQ ID No 255.

XX KW immunogenic; infection; Chlamydia trachomatis; immune; vaccine;
 gene therapy; antibacterial.

XX OS Chlamydia trachomatis.

XX FN WO2003049762-A2.

XX PD 19-JUN-2003.

XX PF 12-DEC-2002; 2002WO-IB0005761.

XX PR 12-DEC-2001; 2001GB-00029732.

XX PR 06-AUG-2002; 2002GB-00018233.

XX PR 14-AUG-2002; 2002GB-00018924.

XX PA (CHIR-) CHIRON SPA.

PI Grandi G, Ratti G;
 DR WPI; 2003-532882/50.
 DR N-PSDB; ADD43961.
 XX
 PT New immunogenic composition having a protein or encoding nucleic acid,
 PT useful for diagnosing, preventing and/or treating Chlamydia trachomatis
 PT infection.
 XX
 PS Claim 6; SEQ ID NO 255; 164pp; English.
 XX
 CC The invention relates to a novel immunogenic composition comprising a
 CC protein or nucleic acid, and an adjuvant, where the protein or nucleic
 CC acid comprises any of 131 fully defined amino acid or nucleotide
 CC sequences given in the specification, or has 50% or greater sequence
 CC identity to it, or their fragments. The protein and/or nucleic acid of
 CC the immunogenic composition is useful in the manufacture of a medicament
 CC for the treatment or prevention of infection due to Chlamydia
 CC trachomatis. The infection is treated or prevented by the medicament
 CC eliciting an immune response which is specific to a C. trachomatis
 CC elementary body, or for neutralising C. trachomatis elementary bodies,
 CC hence the immunogenic composition can be used in creating a vaccine. The
 CC immunogenic compositions can also be used for the diagnosis of C.
 CC trachomatis infection. The nucleic acids of the immunogenic compositions
 CC can be used to treat disorders by gene therapy. The immunogenic
 CC compositions have antibacterial activity. This sequence represents one of
 CC the 131 C. trachomatis proteins with immunogenic properties of the
 CC invention.
 XX
 SQ Sequence 1005 AA;
 Query Match 3.8%; Score 113.5; DB 7; Length 1005;
 Best Local Similarity 21.0%; Pred. No. 0.17;
 Matches 59; Conservative 41; Mismatches 88; Indels 89; Gaps 13

Qy	176	TQTGASEDSG-----DEGSTRHSGAGVQ-PVDDANADSPGS-----GDSG-----	216
Db	491	TQAGPSEDDGHSFNETPGAGAAPSTPSSPIINNVNVGGTINVGIDTNVNTINT	550
Qy	217	-PSTRSDSQPPADETTVHTDNVEDDLTLDKESACA-----LMYHVGQENDMLM	266
Db	551	TPTQTSTASTSIDDINTNNQTDINTTKSDGAGGVNGDISSTESSGDDSGSGVS	610
Qy	267	RAMCED-----LFDLLGIPE--DVIAISQPGDITDA-----SGVYTE	302
Db	611	SSESDKNASGVNDGPGAMKDLISAVRKLHDVVPVCGNGSGTEGPLPANQTLGDV	670
Qy	303	GS-----IAASAVGAGVEDVYLGALEAQNVAAGEVYLEISDEEVDGCA	345
Db	671	GSAQDTKLSGNTGAGDDPTTTAAVNGAEIITLS---DTSIGIGDDVSDTASSSGD	727
Qy	346	GL---PPASRRRPVVGFEFLWDDGP-----RRH	369
Db	728	GVSSPSESNNKTAVG---NDGPSGLDILAARVKH	759

RESULT 11
 ABP96241
 ID ABP96241 standard; protein; 1007 AA.
 AC ABP96241;
 XX
 XX
 DT 19-MAY-2003 (first entry)
 XX
 DE Human nucleic-acid associated protein 24 SEQ ID NO:24.
 XX
 KW Human; nucleic-acid associated protein; NAAP; cardiac; cytostatic;
 KW neuroprotective; gene therapy; cardiovascular disorder; cancer;
 KW neurological disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO2003016549-A2.

QY 379 HRKLSAYRVARPPVMTIDELGVEVFYFGRP-AMSL-VERKVFIL 423
Db 420 FHRREKYPVQMPHPVPEHLDVITSSGLFYGMSPPEKAEBAATPGGVERKPLVA 479
QY 424 CSQNLADISHSCLHSRKG-LRVLLPK-PDDNNTGFG-DVNLLAAV 466
Db 480 STTALSATESLTLLSTAGTATAPGLPAFNKFLVMKAVEPKNKADENTPPGSGSAISGV 539
QY 467 LRSFAS-----GLVIVSLRSGIYVKNLCKST 492
Db 540 AESSTATRMQLSKLVTSLPSWALLTNHFKST 570

RESULT 12
ID ADC06835 standard; protein; 1007 AA.
XX
AC ADC06835;
DT 18-DEC-2003 (first entry)
XX
DE Human prostate cancer-related protein XM_033473.
KW cystostatic; prostate cancer; breast; gene therapy; transgenic; human.
XX
OS Homo sapiens.
XX
PN WO2003064599-A2.
XX
PD 07-AUG-2003.
XX
PP 24-JAN-2003; 2003WO-US001943.
XX
PR 25-JAN-2002; 2002US-00054935.
PR 14-FEB-2002; 2002US-0356130P.
PR 22-MAR-2002; 2002US-00102946.
PR 08-APR-2002; 2002US-00117229.
PR 14-MAY-2002; 2002US-00144198.
PR 19-JUL-2002; 2002US-00197824.
XX
PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;
XX
DR WPI; 2003-679495/64.
XX
PT New isolated polynucleotide related to cancer genes, useful for
PT detecting, diagnosing, staging, monitoring, prognosticating, preventing
PT or treating cancers, e.g. breast and prostate cancers.
XX
PS Disclosure, Fig 18; 128pp; English.
XX
CC The invention relates to a novel isolated polynucleotide comprising a
CC differentially-regulated mammalian cancer gene. The polynucleotides of
CC the invention demonstrate cytostatic activity and are differentially
CC expressed in prostate cancer. The polynucleotide, polypeptides and
CC methods of the invention may be useful for detecting, diagnosing,
CC staging, monitoring, prognosticating, preventing or treating cancers,
CC particularly breast and prostate cancers. Furthermore, the invention may
CC be utilised during gene therapy procedures or in the production of
CC transgenic animals. The current sequence is that of the prostate cancer-
CC related protein of the invention.
XX
SQ Sequence 1007 AA;

Query Match 3.8%; Score 113; DB 7; Length 1007;
Best Local Similarity 19.4%; Pred. No. 0.2;
Matches 111; Conservative 61; Mismatches 171; Indels 228; Gaps 25;
QY 146 KHRPRPR-----PFGVNSDSEEDHFAFCDVPTQ-----TGASED----- 184
Db 4 RKORXPOQLISDCGPGSASNGDASEEDHPQVCAKCAQFTDPTFEFLAHQNACTDPPVM 63

QY 185 ---SGDEGSTRHSASGVQP-----VDDANADSPSGDE-----GPSTRHSDSQ- 225
Db 64 VIIGQENPNNSASSEPRPEGHNNPNQVMDTEHSPNPDSSGVSPTDPTWGPERRGEESG 123
QY 226 -----PPADETTVHTDNVE----- 240
Db 124 HFLVAATGTAAGGGGLILASPKLGATPLPESTAPPPPPPPPGVSGHNAIPLIL 183
QY 241 DDLTLLDK-----SSAC---ALMYHVGQ-----EMDMLMRAMCDEDFDLIG 279
Db 184 EELRVLOQRQIHQMOMTEQICRQVILLGLSGTGVGAPASPSLPGTGTASTKPLPLIFS 243
QY 280 IPEDV-----IATSPGGDT-----DASGV----- 299
Db 244 PIKPVQTSKTLASSSSSSSSSSGASTPKQAFPHLYHPLGSGQHPFAGGVGRSHKTTAPS 303
QY 300 -----VTEGSTAASAVGA--GVEDVYLAGALEAQNVAGEYVLEISD 338
Db 304 PALPGSTQOLIASPHLAPFTTGLLAAQCGLGAARGLEATASGLLKPNKNGS-ELSY 359
QY 339 EEVDDGAGLPPASRRRPVVGELWDGP-----RSH--ERP-----TTR---RIR 378
Db 360 GEVMGPLEKPGGRHKRCFCAKVFGSDSALQIHLRSHTGERPYKNCVGNRFTTRGNLKVH 419
QY 379 HRKLSAYRVARPPVMTIDRLGVEVFYFGRP-AMSL-VERKVFIL 423
Db 420 FHRREKYPVQMPHPVPEHLDVITSSGLFYGMSPPEKAEBAATPGGVERKPLVA 479
QY 424 CSQNLADISHSCLHSRKG-----LRVLLPK-PDDNNTGFG-DVNLLAAV 466
Db 480 STTALSATESLTLLSTAGTATAPGLPAFNKFLVMKAVEPKNKADENTPPGSGSAISGV 539
QY 467 LRSFAS-----GLVIVSLRSGIYVKNLCKST 492
Db 540 AESSTATRMQLSKLVTSLPSWALLTNHFKST 570

RESULT 13
ID AAM79822 standard; protein; 1019 AA.
XX
AC AAM79822;
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3468.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX N-PSDB; AAK52955.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX Claim 20; Page 347-348; 622pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX (AAK52582) and 3686 (AAW80020) are omitted as the relevant pages from the
XX sequence listing were missing at the time of publication
XX Sequence 1019 AA;
SQ
Query Match 3.8%; Score 113; DB 4; Length 1019;
Best Local Similarity 19.4%; Pred. No. 0.2; Indels 228; Gaps 25;
Matches 111; Conservative 61; Mismatches 171; Indels 228; Gaps 25;
QY 146 KRRRPR-----PFVGENSDSEEDHPAFCDVFVTQ-----TGAESD----- 184
Db 16 RKQRKQQLISDCEGFSASENGDAEEDHPQVCAKCAQFTDTEFLAHQACSTDPFVM 75
QY 185 ---SGDEGSTRHSASGVOP-----VDDANADSPGSGDE-----GPSTRHSDQ- 225
Db 76 VIIGQENFNNSASSEPEEGHNPQVMDTSHSPDPSGSSVPTDPTWGPERGESESG 135
QY 226 -----PPADETTVHVDNVE----- 240
Db 136 HFLVAATGTAAGGGGILLASPKLGATPLPPESTAPPPTPPPPPPGSGHLNIPIL 195
QY 241 DLTLLDK-----ESAC---ALMHVQ-----EMDMLRAMCDEDLFDLLG 279
Db 196 BELVLQQRQIHQMOMTEQICRQVLLGLSGTGVGAPAPSPSELPGTGTASSTKPLPLFS 255
QY 280 IPEDV-----IATSPQGGDT-----DASGV----- 299
Db 256 PIKPVQTSKTLASSSSSSSSSGAETPKQAFPHLYHPLGSHQHPFAGGVGRSHKPTAPS 315
QY 300 -----VTEGSTAASVGA--GVEDVYLAGALEAQNVAGEYVLEISD 338
Db 316 PALPGSTDQLIASPHLAFPSITGTLAAQCLGAARGLEATASGLLKPNKNGSG---ELSY 371
QY 339 EEVDDGAGLPPASRRPVPVGEFLMDGP-----RRH---ERP-----TTR---RIR 378
Db 372 GEVMPGLEPGRGKRCFCARFVGSDALSALQIHLRSHGTGERPYKNCVGNRTTRGNLKVH 431
QY 379 HRKLRSAYRVARPPVMTDRLGVEVFFVGRP-AMSL------VERKVFIL 423
Db 432 FHRREKYPHVQMPHPVPEHLDYVITSSGLPYGNSVPEKAEBAEPGGGVKPLVA 491
QY 424 CSQNPLADISHCLSRKG-----LRVLLPK-PDNNVTCGP-DVNLAAV 466
Db 492 STTALSATESLTLTSGTATATAPGLPAFNKFEVLMKAVEPKNKADENTPPGSEGSATSGV 551
QY 467 LRSFAS-----GLIVSLRSAGIYVKNLCKST 492
Db 552 AESSATATRNQSLKVLTSPLSWALLTNHFKST 582

RESULT 14

AD55522
XX ID ADE55522 standard; protein; 466 AA.
XX AC ADE55522;
XX 29-JAN-2004 (first entry)
XX DT
XX XX
XX DE Rat Protein O55170, SEQ ID NO 1339.
XX KW
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX OS Rattus norvegicus.
XX FN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GEHO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX XX
XX DR WPI; 2003-268312/26.
XX DR GENBANK; O55170.
XX PT
XX New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
XX ES Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat
XX or human polynucleotides or a polynucleotide which represents a fragment,
XX derivative or allelic variation of the nucleic acid sequence. Also
XX claimed are a vector comprising the novel polynucleotide, a host cell
XX comprising the vector, a method for identifying a nucleotide sequence
XX which is differentially regulated in an animal subjected to pain and a
XX kit to perform the method, an array, a method for identifying an agent
XX that increases or decreases the expression of the polynucleotide sequence
XX that is differentially expressed in neuronal tissue of a first animal
XX subjected to pain, a method for identifying a compound which regulates
XX the expression of a polynucleotide sequence which is differentially
XX expressed in an animal subjected to pain, a method for identifying a
XX compound that regulates the activity of one or more of the
XX polynucleotides, a method for producing a pharmaceutical composition, a
XX method for identifying a compound or small molecule that regulates the
XX activity in an animal of one or more of the polypeptides given in the
XX specification, a method for identifying a compound useful in treating
XX pain and a pharmaceutical composition comprising the one or more
XX polypeptides or their antibodies. The polynucleotide or the compound that
XX modulates its activity is useful for preparing a medicament for treating
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
XX therapy). The sequence presented is a rat protein (shown in Table 2 of
XX the specification) which is differentially expressed during pain. Note:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic form directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 466 AA;

Query Match 3.8%; Score 112.5; DB 7; Length 466;
Best Local Similarity 20.3%; Pred. No. 0.064;
Matches 83; Conservative 50; Mismatches 143; Indels 133; Gaps 21;
QY 119 WCWHAR-----THLHSG-----SSLWEILYQHSVRLEKRRPRPFVGENSD---S 162
|||||
|||||

Db 114 WAQARRKLDQVPHLHNAELSKTLKWLRLNESD-----KRPTEBAERLRMQ 163
 Qy 163 SEEDHPAFCDVPTQTG-----AESEDSG---DEGPST-----RHASGVOP 201
 Db 164 HKXDPDYKYQPRRRKNGKAAQGEACGGETDQGGAAIQAHYKSAHLDRHPESG-SP 222
 Qy 202 VDDANADSPSGDEGSTRHSDSQPPADSTTV-----HTDNVED 241
 Db 223 MSDGNPEHPSGQSHGPT-----PPTPKTELQSGKADPKRDRSLGEGKPHIDFGNV 276
 Qy 242 DLTLLDKESACALMYHVGQEMDMMLRAMCDEDLFDLLGIPEDVIATSPQ--GUTDASGV 299
 Db 277 DIGEISHE-----VMSNNVEFDVTELDQYLPPNGHPGVGSYGAAY 318
 Qy 300 VTGSIASAAGVAGVEDVYLALAGAEQNVAGEYVLEISDEVDGAGLPPA-SRRPVPVG 358
 Db 319 ---GLSALAVASG---HSAWISKPGVALPTV-----SPPAVDAKAQVKT 358
 Qy 359 EFLWDDGPRRH-ERPTTRIRHRKLR-----SAYYRVARPPVMTDRLGVEYFYFGRPM 412
 Db 359 ETTGPGQPHYTQPSQTSQIAYTSLSLPHYGSAPPSISRPFQDYSQHPGSGP-YIGHAQ 417
 Qy 413 SLEV-----ERKVFILCSQ-NPLADISHSLHSRKLRLVLLPKP 450
 Db 418 ASGLYSAFSYMGPSQRLPYTAISDPSGSPQSHSPHWEQPVVTTLSRP 466

RESULT 15

AD55524
 ID ADE5524 standard; protein; 466 AA.

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

XX AC ADE55524;

CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC XX
 CC Sequence 466 AA;

Query Match 3.8%; Score 112.5; DB 7; Length 466;
 Best Local Similarity 19.5%; Pred No. 0.064;
 Matches 80; Conservative 46; Mismatches 148; Indels 137; Gaps 19;
 QY 119 WCVHAR-----THLHSG-----SSLWEILYQHSVLRKRRRPPFPVGENSD---S 162
 Db 114 WAQARRKLDQVPHLHNAELSKTLKWLRLNESD-----KRPTEBAERLRMQ 163
 QY 163 SEEDHPAFCDVPTQTG-----AESEDSGDEGPST-----RHASGVOP 201
 Db 164 HKXDPDYKYQPRRRKNGKAAQGEACGGETDQGGAAIQAHYKSAHLDRHPESG-SP 222
 QY 202 VDDANADSPSGDEGSTRHSDSQPPADSTTV-----HTDNVED 241
 Db 223 MSDGNPEHPSGQSHGPT-----PPTPKTELQSGKADPKRDRSLGEGKPHIDFGNV 276
 QY 242 DLTLLDKESACALMYHVGQEMDMMLRAMCDEDLFDLLGIPEDVIATSPQGGDTASGVVT 301
 Db 277 DIGEISHE-----VMSNNMETFDVAELDQYLPNGHPG-----H 309
 QY 302 EGSIAASAVGAGVEDVYLALAGAEQNVAGEYVLEISDEVDGAGLPPAS-----RRRPV 356
 Db 310 VSSYSAGYGLG-----SALAVASGSAWISK---PPGVALTVSPGVDAKAQV 356
 QY 357 VGEFLWDDGPRRH-ERPTTRIRHRKLR-----SAYYRVARPPVMTDRLGVEYFYFGRP 410
 Db 357 KTETAGPQGPHTYDQPSQTSQIAYTSLSLPHYGSAPPSISRPFQDYSQHPGSGP-YIGHS 415
 QY 411 AMSLEV-----ERKVFILCSQ-NPLADISHSLHSRKLRLVLLPKP 450
 Db 416 QASGLYSAFSYMGPSQRLPYTAISDPSGSPQSHSPHWEQPVVTTLSRP 466

Search completed: June 9, 2004, 09:01:10
 Job time : 65 secs

CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent

Claim 1; Page; 1017pp; English.

Woolf C, D'urso D, Befort X, Costigan M;
 WPI; 2003-268312/26.
 GENBANK; P56693.

New composition comprising two or more isolated polypeptides, useful for
 preparing a medicament for treating pain in an animal.

Db 361 LWDGPRRHERTTIRIRHRKLSAYRVARPPVMTDRLGVEVYFGRPAMSLVERKV 420
QY 421 FILCSQNLADISHSCLSHSRKGLRVLLPKPDNDNTGPGVGNLLAAVLFSPASGLVIVSLR 480
Db 421 FILCSQNLADISHSCLSHSRKGLRVLLPKPDNDNTGPGVGNLLAAVLFSPASGLVIVSLR 480
QY 481 SGIVYKVLCKSVLVHGNPPKPKFGVICGLSRVLDVFNVAQRIQGHIEHKKTTVFIF 540
Db 481 SGIVYKVLCKSVLVHGNPPKPKFGVICGLSRVLDVFNVAQRIQGHIEHKKTTVFIF 540
QY 541 GDPTSAEQFDMVPLVILKLSRVTCD 567
Db 541 GDPTSAEQFDMVPLVILKLSRVTCD 567

RESULT 2

US-09-252-991A-19246
; Sequence 19246, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19246
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19246

Query Match 4.0%; Score 119.5; DB 4; Length 479;
Best Local Similarity 21.4%; Pred. No. 0.0011;
Matches 93; Conservative 49; Mismatches 163; Indels 129; Gaps 18;
QY 172 DVFVTQTCAESEDSDGEGPSTRHSAGVQVDDANADSPGSGDEGPGSTRHSDSP-----PP 228
Db 3 EVFSTCGSTTRSWRPGDSTR-STSPAMSSASSEANGPGRVKSSTFNDDNGFFMLPP 61
QY 229 ADETTVHTDNVEDDLTLLDKESACALMVHVGQEMDML-----MRAMCDEDFDLILGIPEDV 284
Db 62 LCRRAL-----VGIACRAAYPPAARDMAVASSELHAAADDLPHDLGTGEDP 108
QY 285 I-----ATSPQGGDT-----DASGVTEGSIASAV 310
Db 109 LDAGVHPADRIPLHVAAPAMQLHAFVDHLAQQLGGEQLGFGDGLGQLLLVVVHAAV 168
QY 311 GAGVEDVYLA---GALEAQNVAAGEVYVLEISDEEVD-----GAGLPPASR 352
Db 169 GEDLHVLDLTHFGELEAG-----VLEVRDPAGEALLHVVGDGRQORLPGHRAGV 222
QY 353 RPPVVGELWDGPRRHERPTTRIRHRKLSAYRVARPPVMTDRLGVEVYFGRPAM 412
Db 223 GQALLGEHV-----HQVEEAALGVAQVFLRHPHIVEEQIG-----GVL 261
QY 413 SLEVERKVFILCSQNLADISHSCLSHSRKGLRVLLP-----KPDNDNTGP-----GVNLLAA 465
Db 262 ALHAB-----LLQVAPALEALHAAHQEANGVLVRRIGLRDRDDHQVQQAQVGDHIGAV 316
QY 466 VLRSFASGLVIVSLRSGIYVKNLCKSTVLYHG-----NNPKKFGVICGLSSRAVLD 517
Db 317 -----EQPVVALVRGAAHAQGVAGGRGLGHGEDGLATDDPPQAGLLIGAT---VFG 368
QY 518 VFNVAQRIQ-CHE 530
Db 369 DVPRAQRVQRDHE 382

RESULT 3

US-09-489-039A-13221
; Sequence 13221, Application US/09489039A
; Patent No. 6610838
; GENERAL INFORMATION:
; APPLICANT: Gary Breston et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13221
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13221

Query Match 3.6%; Score 108.5; DB 4; Length 613;
Best Local Similarity 21.3%; Pred. No. 0.025;
Matches 91; Conservative 36; Mismatches 149; Indels 151; Gaps 17;
QY 148 RRRPREP-----FVGENSDSSEEDHPACDVPVTTQTGAESDSGDEGPGSTRHS 195
Db 116 RHRPAQPGRLRSATLLAGRFAG-----YHPFA-----NAENCPPAAGSGGGDRS 159
QY 196 ASGVQVDDANADSPGSGDEGPGSTRHSDSQ---PPPADETTVHTDNVEDDLTLLDKESAC 252
Db 160 AA-----EPGAEQP---EPGQRYRHRPQPRRHPADRPDRHREVR-----198
QY 253 ALMYHVGQEMDMLMRAMCDEDFDLILGIPEDVIATISQPGGDTDASGVVTEGSIASAVGA 312
Db 199 --QRHRRRPHLYR-----PDDAAYGALAG3--GRSSLVAAGHPARHPGPA 241
QY 313 GV--EDVYLAGALEAQNVAAGEVYVLEISDEEVDGAGLPASRRRPPVGFWDGGRHHE 370
Db 242 AVCERDWPAGARRTAGIGE---RLSRAGADLAGAPPAAGRWPI-----APSRRL 291
QY 371 RPTTRIRHRKLSAYRVARPPVMTDRLGVEVYFGRP-----410
Db 292 IASCLLLTAASLLIISLLGAQGFVPLTTIDQVFSALFGDAPRVAMVWNEWRPLRVLMALL 351
QY 411 -AMSLEVERKVFILCSQNLADISHSCLSHSRKGLRVLLPKPDNDNTGPGD-----459
Db 352 IGNALGVSAIFQSILTRNPLGS-----PDVMGFNTGAWSGVLVAMVLF 395
QY 460 VNLLAAVLSFASGLV-----IVSLRSGIYVKNLCKSTVLYHGNPPKFGVICGLSSR 513
Db 396 QNLTAALAAAGGVLTSLVWLLAWRNGIETFL-----IIIGIVR 438
QY 514 AVLDFEN 520
Db 439 AMLVAFN 445

RESULT 4

US-09-347-833-8
; Sequence 8, Application US/09347833
; Patent No. 6294658
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Odell, Joan T.
; TITLE OF INVENTION: Factors Involved in Gene Expression
; FILE REFERENCE: BB-1172
; CURRENT APPLICATION NUMBER: US/09/347,833
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,415
; EARLIER FILING DATE: July 10, 1998
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Microsoft Office 97

; SEQ ID NO 8
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Zea mays
US-09-347-833-8

Query Match 3.6%; Score 107; DB 3; Length 417;
Best Local Similarity 21.3%; Pred. No. 0.019;
Matches 90; Conservative 51; Mismatches 150; Indels 132; Gaps 20;
QY 121 VHARTLHSGSSLWEI-----PACDVPVTQTCAESEDGEGFSTRHSASGVQ-PVDDANADSPG 161
DB 22 VKAKEHDAFMSWQIMSTQKLSRVFMRDSIDLRNKQORRKBGPKKIEVHRD 81
QY 162 SSEDH-----PACDVPVTQTCAESEDGEGFSTRHSASGVQ-PVDDANADSPG 211
DB 82 AAQERHAQSRGLGRGPAVSVV---RRHPMDYCPGPGPSASASSSQSGIRGMPHPSRG 138
QY 212 SGDEGPTRESDDPPADETVHTDNVEDDLTLKESACALMYHVGEMDMLRAMCD 271
DB 139 SQDIRHERIQ-----FDRNVLDPQV-----KDEAITLGPQGLARGMSIR--- 181
QY 272 EDLFDLIGIPEDVIATSQPGDITDASGVVTEGSIASAVGAGVEDVYLAGALEAQNAGE 331
DB 182 -----GQP-----PVSNTPIPSVIDHRRIVSSNG-----YNSAAD 212
QY 332 YVLEISDEE-----VDGAGLPASRRPVVGEFLWDDGPRRHERPTTRIRHR----- 380
DB 213 WTSSGREGDSNRLPDRTSGRIPASSQSASVTS-----QRPASQGRSRKSYSSE 261
QY 381 -KLRS-----AYY--RVAPPPVMTDRLGVEVY-----FGRPAMSLVEVERK 419
DB 262 DELREKSVLTIREYSKADKEVULCIEELNANFVFLVSLWNDSFKERKMERELLAK 321
QY 420 VFI-LGQNFLADISHSLHSRK---GLRVLLPKPDN-NTGPGDYNLLAAVLRSFASGL 474
DB 322 LLVSLCSGR-----HNLKSKQLSDCLSNVLASLEDNLSDAPRATYLGRLIARFVEES 375
QY 475 VIV 477
DB 376 ILL 378

RESULT 5
US-08-072-610-2
; Sequence 2, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,610
; FILING DATE: 19930602
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/07686
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pVMB3.3.1
US-08-072-610-2
Query Match 3.6%; Score 107; DB 1; Length 1018;
Best Local Similarity 22.5%; Pred. No. 0.086;
Matches 66; Conservative 33; Mismatches 92; Indels 102; Gaps 13;
QY 148 RRRPRPFVGENSDSS-----EEDHPAFCD-----VPVTTQGAESDS- 185
DB 362 RRRNRNVEGETEAEAEVSEETPEGEELEATPDDFDALDGTLEETBETAGGETV 421
QY 186 -----GDEGSTRHSASGVQ-----PVDDANADSPGSGDEG----- 216
DB 422 EGEEVTEGEVEGEAEAEVSEETPEGEELEATPDDFDQLEEPSGEGEGEGEALVAVP 481
QY 217 -----PSTRHSOSOP-----PPADETVHTDNVEDDLTLTD----- 247
DB 482 VVAEVEVVVTPAQVKPMVAPTADS-TLFDVLDNDLTYADITSFELPKQLKDPDAGE 540
QY 248 -----KESACALMYHVG---QEMDMLRAMCDEDLFDLLG---IPED---VIATSPG 292
DB 541 AVTVPSKEAPVQVAVGPAQVTEELMQLQEDDFELEGTAEPAGEGELVLEGEGETE 600
QY 293 DTDASGVVTEGSIASAVGAGVEDVYLAGALEAQNAGEVYLE-ISDEVEDDG 344
DB 601 EPREGEPEGEVEPEEELATPDDF-----ELEPTGEVEETVEGEETAEG 648

RESULT 6
US-08-719-822B-2
; Sequence 2, Application US/08719822B
; Patent No. 5874527
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,822B
; FILING DATE: 09/30/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700

```

Query Match          3.6%, Score 107;
Best Local Similarity 22.5%; Pred. No. 0.
Matches 66; Conservative 33; Mismatch 0.

QY 148 RRRPRRPVFGNSDSS-----EDDH-----
Db 362 RRRNRNRVEGETEERAEAGEVESEETPSGESEELL
QY 186 -----GDEGPSTRHSGASGVQ-----PVDDAA-----
Db 422 EGSETVGEERTVEGEAAEGEELEATPEDDPF-----
QY 217 ----PSTRHSDSQP-----PPADETTVHTDNT-----
Db 482 VVAEPVEVVTPAQPVKPMVAPTADE-TLFDVII-----
QY 248 -----KESACALMHHYV--QENDMLNRACDII-----
Db 541 AVTVESKAPQVQPVVAVGPAQVEETELMQLQI-----
QY 293 DTDASGVVTEGSIASAASVAGAGVEDVYIAGALE-----
Db 601 EPREGEETPEEGPEELEATPEDDF-----E-----

```

Query Match 3.5%; Score 105.5;
Best Local Similarity 21.6%; Pred. No. 0.2
Matches 92; Conservative 44; Mismatches
166 DHPAFCDFVPTQTGAESDESDGEGPSTRASAS

592 DHPGQAVVQGLGDAQGAGDQAPAAVQAGGEGGAVAGDPPAGAVVHRAELAQOOR 651
226 PPPADEITVHTD-----NVEDDITLLDKESACALM--YHVQOEMDMRAMCEDIEDLL 278
552 AGRGQAAVVDQAAIIEVGDAGADQDASALVFAFQVGE-----QAL 697
279 GIPEDVIATSPGGDGTASGVVTVTSGSIAASAVGAGVEDVVLGAGALEAQNAGVYVLEISD 338
698 GADPSLLAVVQPGGHQGDAGVAADAAVAAVVQHAGADIHRTLGA---DHAGTAVVEAGA 753
339 EVDVDDGAGLPAS-----RRPVVGE-----FLWDDGPRHREPTTR--IRHR 380
754 LQRHAGIAEQPAALVWGLAGQRTGAGEGATVVOAGARQAFAFADQRAALVVQHA 813
381 KLRSAVYRVARPPVMTDLGVEVYFGRPAMSLVERKVFILCSQNPADISGSLHSR 440
814 AEAHAQAVLAEPTAV-----AVEQF-----AAVQQA-----VAPGQHPLG-LVQQALHGE 859
441 KGLRV-----LLPKPDNNTGPDVNNLLAAV-----LRSFASGLVTVSL 479
860 AQAAVADDLAAAVVQLLAGVHGLRGAG--NLGAVVDLPRLSDAAIRGQPGCLAVVD- 916
480 RSGIYVKNLCKSTVLYHGNNPPKFGVIGLSSRAVLDFVNAVQYRQGHGHIKKTIVFI 539
917 RVGRDLQGLFAD-----QFATLLG-----QAKRRLQ-----VAL 946
540 GGDPTS 545
947 GGDTPS 952

RESULT 9

US-09-621-976-5191
; Sequence 5191, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-5191

Query Match 3.5%; Score 104; DB 4; Length 164;
Best Local Similarity 25.0%; Pred. No. 0.0083;
Matches 38; Conservative 22; Mismatches 62; Indels 30; Gaps 7;

QY 100 QGINCRWRLLCAEVKECW---CVHARTLHSGSLWE-----ILYQHSV--- 142
DB 19 QHVRQCK-----CLFEGH--WYETGKRYLHRSRTAEKALKAKENRLLQOSIGET 72
QY 143 RLEKRRRRPRPFV-----GENSDSSEEDHPAFCDVPVQTGAESDSCDGPSTRHSASG 198
DB 73 NVERKAKKRSKSVTSSSSSSSSASDSSSESESTSTSSSESDDTDESSESSSSSASS 132
QY 199 VQPVDDANADSPGDCGEPST---RHSDSQP 227
DB 133 TSSSSSDSDSSSSSSSSSTSTSSSSDDEPP 164

RESULT 10

US-08-654-482-14
; Sequence 14, Application US/08654482
; Patent No. 6245562
; GENERAL INFORMATION:

APPLICANT: Dalla-Favera, Riccardo
TITLE OF INVENTION: IDENTIFICATION OF GENES ALTERED IN
TITLE OF INVENTION: MULTIPLE MYELOMA
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/654,482
FILING DATE: 28-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 50995
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 451 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-654-482-14

Query Match 3.4%; Score 103; DB 3; Length 451;
Best Local Similarity 22.4%; Pred. No. 0.059; 74; Indels 76; Gaps 12;
Matches 52; Conservative 30; Mismatches 30; Indels 76; Gaps 12;
QY 347 LPPASRRFPVGEFLWDD-----GPRRHERPTTRIRHKLRSAYVR 388
DB 171 MPFLDRS-----WRDVDPDPHEIPYQCPMTFGRGHHWQGCACNGCQVGTFFA 222
QY 389 VARP-----PVMITD-----RLGVEVYFGRPAMSLVERKVF--LCSQNP 428
DB 223 CAPPEQAFQVPTPEPSIRSAEALAFSDCLRHICLYY-----REILVKELTTSP 271
QY 429 L-ADISHSCLHSRKL-RVLLPKPDNNTGPDVNNLLAAVLRFSFASGLVIVSLRSGIYVK 486
DB 272 EGRISHGHYTDASNLQDLVFPYEDNGHRKNIELLSHLR-----GVVLANPADGLYAK 327
QY 487 NLCKSTVLYHG-----NNPPKFGVICGLSSRAVLDFVNAVQY--RIQHEH 531
DB 328 RLCQSTIYWDGPLALCNDPRNK-----LERDQTKLFDTQQLFLSELQAFAH 373

RESULT 11

US-09-252-991A-23689
; Sequence 23689, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23689

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (713) 789-2679
 TELEX: 79-0924
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1085 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-08-938-534-28

Query Match 3.4%; Score 102; DB 2; Length 1085;
 Best Local Similarity 18.9%; Pred. No. 0.33;
 Matches 60; Conservative 41; Mismatches 115; Indels 102; Gaps 11;
 QY 145 EKRRRRPRPF-----VGENSSSEEDHPAFCDVPVVTQTGAESDSGDEGPSTRH 194
 Db 468 EQERKQRKLYKTKPSTRTTNSVDNDEYFNVFFQ-----SDDENSCHKSKGRH 520
 QY 195 SASGVQPVDDANADS---PGSGDEGPSTR-----HSDSQPP 227
 Db 521 K-SGSHIEHKGNLKNDDLEPSTHSTVLSNGKYDSDDEYDNLDDVAHMFSDDE 579
 QY 228 PADETTVHTDNVEDDLTLIDKESACALMYHVGQEMDMLMRAMCDLDFLLGIPEDVIAT 287
 Db 580 CSESETSHDADTDELRALSDS-----LDIGTEL-----DDYED----- 615
 QY 288 SPOGSDTASGVV-----TEGSIAASAVGAGVEDVYLAGALSAQNV 328
 Db 616 -----DDDDSSVTNVFIDLDLPDSFYHYDSGDS--SSLISNSDKENSCKDCKHD 668
 QY 329 AGEVYLEISDEEVDGAGLPASRRRPVVGFEFLWDGPRRHERPTTTRIRHRKLSAYR 388
 Db 669 LLETVVVDDESTDENLPPSSRSKNIGS-----KAKEIVSSNVVGLRPPKLGWE 721
 QY 389 VARPPVMTDRLGVEVFY 406
 Db 722 TDNKPFSIIDGLSTKSLY 739

RESULT 14

US-09-345-294-28
 Sequence 28, Application US/09345294
 Patent No. 6387619
 GENERAL INFORMATION:
 APPLICANT: Gottschling, Daniel E.
 SINGER, Miriam S.
 TITLE OF INVENTION: Telomerase Compositions and Methods
 NUMBER OF SEQUENCES: 32
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: TEXAS
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/345,294
 FILING DATE: 30-Jun-1999
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/431,080
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Parker, David L.
 REGISTRATION NUMBER: 32,165
 REFERENCE/DOCKET NUMBER: ARCD:155/PAR

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (713) 789-2679
 TELEX: 79-0924
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1085 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 28:
 US-09-345-294-28

Query Match 3.4%; Score 102; DB 4; Length 1085;
 Best Local Similarity 18.9%; Pred. No. 0.33;
 Matches 60; Conservative 41; Mismatches 115; Indels 102; Gaps 11;
 QY 145 EKRRRRPRPF-----VGENSSSEEDHPAFCDVPVVTQTGAESDSGDEGPSTRH 194
 Db 468 EQERKQRKLYKTKPSTRTTNSVDNDEYFNVFFQ-----SDDENSCHKSKGRH 520
 QY 195 SASGVQPVDDANADS---PGSGDEGPSTR-----HSDSQPP 227
 Db 521 K-SGSHIEHKGNLKNDDLEPSTHSTVLSNGKYDSDDEYDNLDDVAHMFSDDE 579
 QY 228 PADETTVHTDNVEDDLTLIDKESACALMYHVGQEMDMLMRAMCDLDFLLGIPEDVIAT 287
 Db 580 CSESETSHDADTDELRALSDS-----LDIGTEL-----DDYED----- 615
 QY 288 SPOGSDTASGVV-----TEGSIAASAVGAGVEDVYLAGALSAQNV 328
 Db 616 -----DDDDSSVTNVFIDLDLPDSFYHYDSGDS--SSLISNSDKENSCKDCKHD 668
 QY 329 AGEVYLEISDEEVDGAGLPASRRRPVVGFEFLWDGPRRHERPTTTRIRHRKLSAYR 388
 Db 669 LLETVVVDDESTDENLPPSSRSKNIGS-----KAKEIVSSNVVGLRPPKLGWE 721
 QY 389 VARPPVMTDRLGVEVFY 406
 Db 722 TDNKPFSIIDGLSTKSLY 739

RESULT 15

US-09-252-991A-32880
 Sequence 32880, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 32880
 LENGTH: 1224
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-32880
 Query Match 3.4%; Score 101.5; DB 4; Length 1224;
 Best Local Similarity 22.6%; Pred. No. 0.46;
 Matches 93; Conservative 30; Mismatches 166; Indels 123; Gaps 18;
 QY 143 RLEKRRRRPRPFVGENSSSEEDHPAFCDVPVVTQTGAESDSGDEG----- 189
 Db 58 RAQRDRSRHRPGRVRRRLRQAAEPA-----RTGVAGEESDPPGAPCQRRPRPAG 110
 QY 190 -PSTRHSAGVQPVDDANADSPGSGDEGFS-----TRHSDSQ 225

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 09:02:41 ; Search time 50 seconds
(without alignments)
3190.381 Million cell updates/sec

Title: US-10-656-499-2

Perfect score: 2993

Sequence: 1 MAGRRITWSEFVIGALDSD.....QFDMVPLVIKLRVSTCDD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pcp.*
- 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pcp.*
- 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pcp.*
- 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pcp.*
- 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pcp.*
- 6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
- 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pcp.*
- 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pcp.*
- 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pcp.*
- 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pcp.*
- 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pcp.*
- 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pcp.*
- 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pcp.*
- 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pcp.*
- 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pcp.*
- 16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pcp.*
- 17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pcp.*
- 18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121.5	4.1	2517	12	US-10-087-192-654
2	119	4.0	691	12	US-10-425-114-45486
3	115	3.8	2507	10	US-09-819-104A-2
4	114	3.8	691	15	US-10-108-260A-4345
5	113	3.8	1007	16	US-10-197-824-38
6	112	3.7	1007	15	US-10-144-198-35
7	111	3.7	509	12	US-10-282-122A-61933
8	111	3.7	1216	15	US-10-028-248A-72
9	111	3.7	1216	15	US-10-107-782-72
10	108	3.6	810	12	US-10-425-114-40809
11	108	3.6	953	12	US-10-282-122A-64713
12	107	3.6	883	10	US-09-759-130B-342
13	107	3.6	883	14	US-10-189-123-72
14	107	3.6	883	14	US-10-188-495-72
15	107	3.6	1134	12	US-10-425-114-45409

Sequence 27, Appl
Sequence 17, Appl
Sequence 104, App
Sequence 54050, A
Sequence 63586, A
Sequence 554, App
Sequence 34, Appl
Sequence 2, Appl
Sequence 7, Appl
Sequence 2342, Ap
Sequence 854, App
Sequence 1510, Ap
Sequence 1400, Ap
Sequence 41846, A
Sequence 3031, Ap
Sequence 8031, Ap
Sequence 13136, A
Sequence 1401, Ap
Sequence 254370, A
Sequence 62763, A
Sequence 64499, A
Sequence 2, Appl
Sequence 62, Appl
Sequence 14, Appl
Sequence 60357, A
Sequence 46877, A
Sequence 13, Appl
Sequence 52062, A
Sequence 3305, Ap
Sequence 65986, A

US-10-297-022-27
US-10-384-919-17
US-10-072-012-104
US-10-425-114-54050
US-10-425-114-63586
US-10-307-817-554
US-10-197-824-34
US-10-346-863-2
US-10-197-824-7
US-10-264-049-2942
US-09-764-864-854
US-10-412-699B-1510
US-10-374-780A-1400
US-10-425-114-41846
US-10-128-714-3031
US-10-128-714-8031
US-10-369-493-13138
US-10-374-780A-1401
US-10-424-599-254370
US-10-282-122A-62763
US-10-282-122A-64499
US-10-384-919-2
US-09-964-956-62
US-10-120-801-14
US-10-425-114-60357
US-10-425-114-46877
US-09-291-417-13
US-10-425-114-52062
US-10-128-714-3305
US-10-425-114-65986

ALIGNMENTS

RESULT 1

US-10-087-192-654
; Sequence 654, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 654
; LENGTH: 2517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-192-654

Query Match 4.1%; Score 121.5; DB 12; Length 2517;
Best Local Similarity 21.3%; Pred. No. 0.28;
Matches 81; Conservative 48; Mismatches 132; Indels 119; Gaps 17;
QY 105 RWELLCAEV--KECWCVHARTLHSGSSILWELLYOHSVRLKHKRRPRPFVGENSDS 162
DB 632 RNWAIARMGSKTVSOCKNFYFKYKQNLDELLOQHKLMKERNARKKKAPAAAS 691
QY 163 SEEDHPAFCDVPVTQ-----TGAESE-----DSGDE-----GFSRHSAS 197
DB 692 EEAFAFP-----PVVEDEMEASGVSGNEERMEVEAEALHASGNEVPRGCGPATVNSS 746
QY 198 GVQ-----PVDANADSPGSGDEGPSTHSDSQP-----PADETT--- 233

Db 540 AESSTATRMQLSKLVTSLPSWALLTNHFKST 570

RESULT 7

US-10-282-122A-61933
; Sequence 61933, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 61933
; LENGTH: 509
; TYPE: PRF
; ORGANISM: Mycobacterium avium
US-10-282-122A-61933

Query Match 3.7%; Score 111; DB 12; Length 509;
Best Local Similarity 24.5%; Pred. No. 0.24;
Matches 54; Conservative 31; Mismatches 65; Indels 70; Gaps 12;
QY 172 DVPVQTGAEEDESGDEGSPRHSAGVQPYD-----DANADSPGSGDEGSPRHS- 222
Db 11 DGPVTRATKSPSPAKPAK-NANGSAPAKRATKTASRSKSEACAAEPKATRSSA 69
QY 223 ---DSQPP-----PAETTVHTDNVEDDL-TLLDKESACALMYHVGQEMDLMR 267
Db 70 KGADAKAPSGRGTKAAGKGPADPALDTGAVEDLDTEPDLEG-----EPGEDLDI--- 120
QY 268 AMCEDDFDLIGIPEDVIATIS---QPGGDDTDASGVVTEGSTAASVACGVEDVYLAGALE 324
Db 121 ---DTDL-NLDDLEEDVAADADIEPG---DAEAGEDEEAAPKAAGATAAD----- 165
QY 325 AQNVAGEYVLEISDEEVDGAGLPAPRRFPVVGFEFLWDD 364
Db 166 -----EDEIAEPSEKDKAS-----GDFVWDE 187

RESULT 8

US-10-028-248A-72
; Sequence 72, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
; LENGTH: 1216
; TYPE: PRF
; ORGANISM: Rattus norvegicus
US-10-028-248A-72

Query Match 3.7%; Score 111; DB 15; Length 1216;
Best Local Similarity 20.4%; Pred. No. 0.91;
Matches 99; Conservative 49; Mismatches 154; Indels 184; Gaps 25;
QY 30 RTGTFPLAARNDVPLDSLOFFIDFKRCLSKGLHPRDLGSPITAFGKICTSRRLR 89
Db 708 RQGGALLTRPQPVVP-----KECLSFMYH---LHGPQI---GTLCLAMRR-- 748
QY 90 RLPGGEYEVW---QGINCRWRLLCAEVKEQWCVCVHARTLHSGSSLWEIL-----YQ 139
Db 749 ---EGEDILLNSRSGTHGNRWI-----QAWVTLH---HQLPSTKYQLLFEGLRDCYH 796
QY 140 HSVRLKHKRRPRRPFVGENSDSSEEDHPACDVPVVTGTGAESDSDGEGGSTR-HSASG 198
Db 797 GTMGLDMAVRPGPCWAAKRCFSFDS---CGF-----SPGDWGLWTRQNNASG 842
QY 199 VQP-----VDDANADSPGS---GDEGP-----STRHSDSQPPPADETTVHTDNVEDDL 243

Db 843 LGPWGFWIDHTTGAQGHVWVDTSPNLLPKGHVASLTSEHPP-----886
QY 244 TLLDKESACALMYHVGQMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTASGVVT--301
Db 887 --LSRPACLSPWHLSPHNPCTLRVFE-----STRQELSIHGFGFAWRLGSVNVQ 938
QY 302 ---EGSIAASAVGAGVEDVYLAGALEAQNVAGYVLEISDEEYDDGAGLPPAS-----351
Db 939 AEQAWKVVFEMASGVHSYMA-----LDDISLQDGFCAQPGSCDFESGL 983
QY 352 -----RRRPVVGFLND-----DGPRHERPTTRIRHKLRSAYYVARPPVMTDRLG 402
Db 984 CGWSHLPWPLGGYSDWSSGATPSRYPRPS-----VDHTVGT 1021
QY 403 EVFYF-----GRPAMSLEVERKVFLICSONPLADISHSCL-----HSRKG 442
Db 1022 EAGHPAFFETSVLGGQQAALGSE-----PLPATAVSCLHFWYMGFFAIFYKG 1071
QY 443 -LRVLL 447
Db 1072 ELRVLL 1077

RESULT 9

US-10-107-782-72
; Sequence 72, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Bojdog, Ferenc
; APPLICANT: Casman, Steve
; APPLICANT: Colman, Steve
; APPLICANT: Edinger, Shomit,
; APPLICANT: Gangolli, Bsha,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Li, Li,
; APPLICANT: Liu, Xiaochong,
; APPLICANT: Malyankar, Uriel,
; APPLICANT: Miller, Charles,
; APPLICANT: Millet, Isabelle,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Sciore, Paul,
; APPLICANT: Shenoy, Suresh,
; APPLICANT: Shinkets, Richard,
; APPLICANT: Si, Jingsheng,
; APPLICANT: Smithson, Glennda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Stone, David,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28

; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: Curaseq1 version 0.1
; SEQ ID NO 72
; LENGTH: 1216
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-72

Query Match

Best Local Similarity 3.7%; Score 111; DB 15; Length 1216;
Matches 99; Conservative 49; Mismatches 154; Indels 184; Gaps 25;

QY 30 RSTGTFLAPARNDIVPLDSLOFFIDFKRECLSGKLHPRDLGSPITAFGKICTTSRLR 89
Db 708 RQGGALLTRPQVPVP-----KECLSPWYH---LHGPOI---GTLCLAMRR-- 748
QY 90 RLPGEYEW---QGINCRWRLLCAEVKECWCWCVHARTHLHSGSSILWEIL-----YQ 139
Db 749 --EGEEDTLLNSRSGTHGRWH-----QAWTLH---HQLPSTKYQLFEGLRDGYH 796
QY 140 HSVRLKHHRRPRPRPFVGENSDSSEEDHPAFCDVPVQTGAESDSDGDEGSTR-HSASG 198
Db 797 GTMGDDMAVRPGPCWAARKCSFEDSD---CGF-----SPGDWGLWTRQNNASG 842
QY 199 VQP-----VDDANADSPGS---GDEGP-----STRHSDSQPPPADETTVHTDNVEDDL 243
Db 843 LGPWGFWIDHTTGAQGHVWVDTSPNLLPKGHVASLTSEHPP-----886
QY 244 TLLDKESACALMYHVGQMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTASGVVT--301
Db 887 --LSRPACLSPWHLSPHNPCTLRVFE-----STRQELSIHGFGFAWRLGSVNVQ 938
QY 302 ---EGSIAASAVGAGVEDVYLAGALEAQNVAGYVLEISDEEYDDGAGLPPAS-----351
Db 939 AEQAWKVVFEMASGVHSYMA-----LDDISLQDGFCAQPGSCDFESGL 983
QY 352 -----RRRPVVGFLND-----DGPRHERPTTRIRHKLRSAYYVARPPVMTDRLG 402
Db 984 CGWSHLPWPLGGYSDWSSGATPSRYPRPS-----VDHTVGT 1021
QY 403 EVFYF-----GRPAMSLEVERKVFLICSONPLADISHSCL-----HSRKG 442
Db 1022 EAGHPAFFETSVLGGQQAALGSE-----PLPATAVSCLHFWYMGFFAIFYKG 1071
QY 443 -LRVLL 447
Db 1072 ELRVLL 1077

RESULT 10

US-10-425-114-40809
; Sequence 40809, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingtong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 40809
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Schizochytrium aggregatum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3033-046-E6_FLI.pep
US-10-425-114-40809

Query Match 3.6%; Score 108; DB 12; Length 810;
Best Local Similarity 25.2%; Pred. No. 0.94; Indels 62; Gaps 10;
Matches 62; Conservative 25; Mismatches 25; Indels 62; Gaps 10;
QY 179 GAESDSDGSPTRHSGVQVVDNADSPGSDGSPTRHSDSQPPADTTVHTDN 238
DB 467 GSRGSGGRRGRATSDLGHGIDNL-ANSAGAA-EGIDVLGLDVE-----HGLN 515
QY 239 VEDDLTLDKESA-CALMYHVQEMDMLRAMCDEDLF-DLLGIPEDVIATSPQGDGTD 296
DB 516 LGEDLDTLDGVNALSLEHLVGLDHSVLASLGEDLHDSLSFLEKLVDRDSGGSGRR 575
QY 297 SGVVTGSIASAVGAGVEDVYL-----AGALEAQNAG----- 330
DB 576 SGSRSSSGRRRGGAASDLSLHGIDLTNSASAESVDVLGVEHGLNLGENDLTLD 635
QY 331 -----EYLETISDEV-DGAG-----LPPASRRPVVGEFLWDGPPR 368
DB 636 GVNAGLGLHVLHVLHVSLSVSLGSGEDLHDSLSLLEKLARDRRRGSSR-TWGGGRR 694
QY 369 HERPTT 374
DB 695 RGRATS 700

RESULT 11
US-10-282-122A-64713
; Sequence 64713, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITPA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 64713
; LENGTH: 953
; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64713
Query Match 3.6%; Score 108; DB 12; Length 953;
Best Local Similarity 22.0%; Pred. No. 1.2; Indels 160; Gaps 24;
Matches 106; Conservative 50; Mismatches 166; Indels 160; Gaps 24;
QY 83 TTSRR-----LRLPQ-----EYEVVQGINCREWRLLCAEVEKWCVHARTLHSGSSLWE 135
DB 34 TTSRRVLTALDGRVRSASHTVDRDAVRDLLA-----THLETAGVL-- 79
QY 136 ILYQHSVLEKRRRPRPPFVGENSDSEED-----HPAF-CDVPVTTQCAESDSDGE 188
DB 80 ---AASVHAPEASEEPESRLMLETQETRNADVPERHYMPLEFVAPQPIPEPLADDEDV-DD 135
QY 189 GPSTRHSASGVQVVDNADSPGSDGEGPSTRH-----SDSQP----- 226
DB 136 GPD-----YVADDSADDEGQLDRPANRRRRRRGRGRGRGEGGSDGDPVDOQSE 187
QY 227 PPA-----DETTVHTDNNVEDDLTLLD-----KESACALMYH 257
DB 188 PRAQOFTSADAAETDDGDRDSEDTEAGDNGEDENGSLAENRRRRRRRRRSKAS----- 242
QY 258 VQEMDMLMRAVCDEDLFDLLGIPEDVIATSPQGDGTDASGVVTEGSLAASAVGAGVEDV 317
DB 243 -GDDNDAALSGPLPDD-----PPNTVHVHVFRAGDKAGNSQDGG----- 281
QY 318 YLGALEAQNAGVEXYLEISDEVDGAGLPPASRRPVV---GEFLWDDGPRRH--BRP 372
DB 282 --SGSTEIKGIDGSTRLEAKRQRDRGR---DAGRRRPPVLSEAEFL---ARREAVRV 332
QY 373 TTRIRHKLRSAYYRVARPVV-----WITDRLGVEVYFGRPAMSLVERKVFIL 423
DB 333 MVDRVR-----TEPLPGTRYTQIAVLE--GIVVEHFTVTSASASLVGNIVILG 381
QY 424 CSQNPFLADISHSLSRKGLRVLLPKDDNNTGPDVNLAAVLRSFASGLVIVSLRSGI 483
DB 382 IVQNVLPSEAAAFVDIGRNGVL-----YAGEVNWDAAGLGG-ADRKTQALKPGD 432
QY 484 YV 485
DB 433 YV 434
RESULT 12
US-09-759-130B-342
; Sequence 342, Application US/09759130B
; Publication No. US2003002279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirst, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: MPI00-5350WNIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24

; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/508,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 342
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-759-130B-342

Query Match 3.6%; Score 107; DB 10; Length 883;
Best Local Similarity 24.0%; Pred. No. 1.3;
Matches 78; Conservative 19; Mismatches 104; Indels 124; Gaps 16;
QY 151 PRPFTVGEN-----SDSSEEDHPA-----PCDVPVTQTGAESDSC-----DE 188
Db 423 PRTPLESETQSIAPPTESSEEGVALEEEERFKDLEALEEKEQEDLWVWPRELSSPLPT 482
QY 189 GPSTRHSASGVQVDDANADSPGSGDEGSTRHSQPPPADETTVHTDNVEDDLTLDDK 248
Db 483 GSETEHLSQVSPPAQAVQLDASPSGPPR-----FRGPPAE-----TLLP- 524
QY 249 ESACALMYHVGQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTDASGVVTEGSIAS 308
Db 525 -----PREMSATSTPGG-----AR 538
QY 309 AVGAGVEDVYLAGALEAQNVAAGEYVLEISDEEVDGAGLPPASRRRPVVGEBFLWDD-GPR 367
Db 539 EVGGETGSPELSGVPRESEEGSSLE-----DGPSSLPAT-----WAPVGR 581
QY 368 RHERPTTRIRHKLRSAYRYVARPPVMTDRL-----GVEVFYFGRPAMSLVERKVFILC 424
Db 582 ELETPSEKSGRTVL--AGTSVQAQVLPDTSASHGVAV-----APSSG-----DC 626
QY 425 SONPLADISHSCLHSKRGRLVL-LP 448
Db 627 IPSPCHN-GGTCLEKEGFRCLCLP 650

RESULT 13
US-10-189-123-72
; Sequence 72, Application US/10189123
; Publication No. US20030082586A1
; GENERAL INFORMATION:
; APPLICANT: KIRST, Susan J.
; APPLICANT: HOLTZMAN, Douglas A.
; APPLICANT: FRASER, Christopher C.
; APPLICANT: SHARP, John D.
; APPLICANT: BARNES, Thomas S.
; TITLE OF INVENTION: ANTIBODIES HAVING DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: 10147-11U3
; CURRENT APPLICATION NUMBER: US/10/189,123
; PRIOR FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 72
; LENGTH: 883

; TYPE: PRT
; ORGANISM: Mus sp.
US-10-189-123-72
Query Match 3.6%; Score 107; DB 14; Length 883;
Best Local Similarity 24.0%; Pred. No. 1.3;
Matches 78; Conservative 19; Mismatches 104; Indels 124; Gaps 16;
QY 151 PRPFTVGEN-----SDSSEEDHPA-----PCDVPVTQTGAESDSC-----DE 188
Db 423 PRTPLESETQSIAPPTESSEEGVALEEEERFKDLEALEEKEQEDLWVWPRELSSPLPT 482
QY 189 GPSTRHSASGVQVDDANADSPGSGDEGSTRHSQPPPADETTVHTDNVEDDLTLDDK 248
Db 483 GSETEHLSQVSPPAQAVQLDASPSGPPR-----FRGPPAE-----TLLP- 524
QY 249 ESACALMYHVGQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTDASGVVTEGSIAS 308
Db 525 -----PREMSATSTPGG-----AR 538
QY 309 AVGAGVEDVYLAGALEAQNVAAGEYVLEISDEEVDGAGLPPASRRRPVVGEBFLWDD-GPR 367
Db 539 EVGGETGSPELSGVPRESEEGSSLE-----DGPSSLPAT-----WAPVGR 581
QY 368 RHERPTTRIRHKLRSAYRYVARPPVMTDRL-----GVEVFYFGRPAMSLVERKVFILC 424
Db 582 ELETPSEKSGRTVL--AGTSVQAQVLPDTSASHGVAV-----APSSG-----DC 626
QY 425 SONPLADISHSCLHSKRGRLVL-LP 448
Db 627 IPSPCHN-GGTCLEKEGFRCLCLP 650

RESULT 14
US-10-188-495-72
; Sequence 72, Application US/10188495
; Publication No. US2003017573A1
; GENERAL INFORMATION:
; APPLICANT: KIRST, Susan J.
; APPLICANT: HOLTZMAN, Douglas A.
; APPLICANT: FRASER, Christopher C.
; APPLICANT: SHARP, John D.
; APPLICANT: BARNES, Thomas S.
; TITLE OF INVENTION: POLYPEPTIDES HAVING DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: 10147-11U2
; CURRENT APPLICATION NUMBER: US/10/188,495
; PRIOR FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 72
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-188-495-72

Query Match 3.6%; Score 107; DB 14; Length 883;
Best Local Similarity 24.0%; Pred. No. 1.3;
Matches 78; Conservative 19; Mismatches 104; Indels 124; Gaps 16;
QY 151 PRPFTVGEN-----SDSSEEDHPA-----PCDVPVTQTGAESDSC-----DE 188
Db 423 PRTPLESETQSIAPPTESSEEGVALEEEERFKDLEALEEKEQEDLWVWPRELSSPLPT 482
QY 189 GPSTRHSASGVQVDDANADSPGSGDEGSTRHSQPPPADETTVHTDNVEDDLTLDDK 248
Db 483 GSETEHLSQVSPPAQAVQLDASPSGPPR-----FRGPPAE-----TLLP- 524
QY 249 ESACALMYHVGQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTDASGVVTEGSIAS 308

Db 525 -----PRENSATSTPGG-----AR 538
Qy 309 AVGAGVEDVVLAGEAONVAGEVLEISDEEVDGAGLPPASRRPVPVGEFLWDD-GPR 367
Db 539 EVGGETGSPGLGVPRESSEAGSSLE-----DGPSLLPAT-----WAPVGPGR 581
Qy 368 RHERPTTTRIRHRKLRGAYRVARPPVMTDRL--GVEVFYFGRPAMSLVERKVFIIC 424
Db 582 ELETSPSEKSGRTVL--ACTSVQAQVLPDTSASHGVAV-----APSSG-----DC 626
Qy 425 SONPLADISHSLHSRKLRLV-LP 448
Db 627 IPSCHN-GGTCLBEKEGFRCLCLP 650

RESULT 15
US-10-425-114-45409
; Sequence 45409, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Jingdong
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45409
; LENGTH: 1134
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700152164_FLI.pep
US-10-425-114-45409

Query Match 3.6%; Score 107; DB 12; Length 1134;
Best Local Similarity 21.3%; Pred. No. 2;
Matches 90; Conservative 51; Mismatches 150; Indels 132; Gaps 20;
Qy 121 VHARTLHSGSSSLWEI-----LYQHSVLEKRRPRPPFVGE-----NSD 161
Db 706 VKAKEHMDAYFMMQIMSTNQKLSRVRFMLRDSIDLRRNKQORRKVEGPKIEVHRD 765
Qy 162 SSERDH-----PACDVPVTQCAESESDGEGPSTRHSASGVQ-PVDDANADSPG 211
Db 766 AAQERHAQSSRLGCGPAVSSVP---RRAHPMDYGPGRPSASASSSQSGSIRGNPHPSRG 822
Qy 212 SGDEGPSTRHSDSQPPADETTVHTDNVEDDLTLDKESACALMYHVGQEMDMLEAMCD 271
Db 823 SQDIRHDERHQ-----FDNRTVLPQVW-----KDEAITLGPQGLARGMSIR---- 865
Qy 272 EDLPDLGIPEDVIATSQPGDITDASGVTEGSIASAVGAGVEDVYLAGALEAQNVAQE 331
Db 866 -----GQP-----PVSNTPEIPSDHRRIVSSNG-----YNSAAD 896
Qy 332 YVLEISDEE-----VDDGAGLPASRRRPVVGFEFLWDGPPRRHERPTTTRIRHR----- 380
Db 897 WTSSSGREDSNRLPDRTSGRIPASSQSAVTS-----QRPASQEGRSRSKSYSE 945
Qy 381 -KLRS-----AYY--RVARPPVMTDRLGVEVY-----FGPAMSLVERK 419
Db 946 DELREKSVLTIREYISAKDEKVVLCIEELNAPFPFLVSLWVNDSPFERKOMERELLAK 1005
Qy 420 VFI--LCSONPLADISHSLHSRK---GLRVLLPKPDN-NTGPGDVNLLAAVLRSPASGL 474
Db 1006 LLYSLCSGR-----HNLSSKQQLSDGLSNVLASLEDNLSAPRATYELGRLLARFVES 1059
Qy 475 VIV 477

Db 1060 ILL 1062
Search completed: June 9, 2004, 09:08:40
Job time : 52 secs

Qy	522	----	AQYRIQGH-E-HIKKTTVFIGDDPSAQGFOWPLVIKLRLSVTCDD	567
		:	:	:
Db	307	LRKLALSPVGGDPVFINAVTLVLRGPGSRK-RQVFVIL-----VICOD	350	
		:	:	:

RESULT 2

```

US-10-779-597-105
; Sequence 105, Application US/10779597
; GENERAL INFORMATION:
; APPLICANT: Oregon Health & Science University
; APPLICANT: Wong, Scott W.
; APPLICANT: Axelholm, Michael K.
; APPLICANT: Hansen, Scott G.
; TITLE OF INVENTION: JAPANESE MACAQUE HERPESVIRUS NUCLEIC ACID SEQUENCES AND THEIR USE
; FILE REFERENCE: 178-67426
; CURRENT APPLICATION NUMBER: US/10/779,597
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/276,524
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: PCT/US01/16274
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/205,652
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 105
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Japanese Macaque Herpesvirus
US-10-779-597-105

```

[illegible]

RESULT 3

US-10-021-698A-122
 : Sequence 122, Application US/10021698A
 : GENERAL INFORMATION:
 : APPLICANT: KEITH, TIM
 : APPLICANT: LITTLE, RANDALL
 : APPLICANT: VAN EERDEWEGH, PAUL
 : APPLICANT: DUPUIS, JOSSE
 : APPLICANT: DEL MASTRO, RICHARD
 : APPLICANT: SIMON, JASON
 : APPLICANT: ALLEN, KRISTINA
 : APPLICANT: PANDIT, SUNIL
 : TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
 : TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
 : FILE REFERENCE: 2976-404AUS1
 : CURRENT APPLICATION NUMBER: US/10/021, 698A
 : CURRENT FILING DATE: 2001-10-22
 : PRIOR APPLICATION NUMBER: 60/211,749
 : PRIOR FILING DATE: 2000-06-14
 : NUMBER OF SEQ ID NOS: 6160

```

; SOFTWARE: Patentin 2.1
; SEQ ID NO 122
; LENGTH: 2517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-698A-122

```

Query Match	4.1%;	Score 121.5;	DB 6;	Length 2517;
Best Local Similarity	21.3%;	Pred. No. 1.4;		
Matches	81;	Conservative 48;	Mismatches 132;	Indels 119; Gaps 17;
QY 105	RRWELLCAEV--KECWCVHARHLHSGSLWEILYQHVSRLKEKRRRRPRPFVGENSDS	162		
Db 532	RNWSAARVMGSKTVSQCKNFYNYKRONLDEILQOHKLKWEKERNARKKKKAPAAAS	691		
QY 163	SEEDHFAFCVPTVQ-----TGAESE-----DSGDE-----GPSTRHAS	197		
Db 692	EEAAFP-----PVVEDEEMEAAGVSGNEBWEAEALHASGNEVPRGCGSPATVNSS	746		
QY 198	GVQ-----PYDDANADSPGSGDEGSPSTRHSDSQP-----PADETT---	233		
Db 747	DTESIFSPHTEAAKQTQGNQPKPATLGADGGPPGPTPPRSTRPIETPTPASEATGAP	806		
QY 234	-----YHTDNVEDDLTLLDKESACALMYHVQEMDLMRAMCDEDLFDLL	278		
Db 807	TPPPAPPSAPPVVPVKEEK-----ETAAAPVVEGEE-----QKPAAEELAVDT	855		
QY 279	GLPDEDVIAIS-----QPGDDTASGVVTGEGSIAASAVGAGVEDVYIAGALEAQNVA	329		
Db 856	GRAEFPVKSECTEEAEEGPAKGDAEAAEATAGALKAEKKEGG-----SGRATTAKSS	909		
QY 330	G-----EYVLETSDEBVDGAG-----LPPASRRRPVVGFEFLDGDGRHRHP-ITTRI	377		
Db 910	GAPQDSDSATCSADBDVBAEGDGKURLSPRPSLLTPTGD-----PRANASQPKPLD	963		
QY 378	RHRKLRSAYYRVARPVPMIT	397		
Db 964	KOLKORAA-----AIPPIQVT	979		

RESULT 4

```

US-60-550-051-213
; Sequence 213, Application US/60550051
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001508
; CURRENT APPLICATION NUMBER: US/60/550,051
; CURRENT FILING DATE: 2004-03-05
; NUMBER OF SEQ ID NOS: 23014
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 213
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-550-051-213

```

```

Query Match      3.9%; Score 117; DB 7; Length 1261;
Best Local Similarity 19.7%; Pred. NO. 1.2;
Matches 122; Conservative 55; Mismatches 187; Indels 256; Gaps 29;

QY      24 LVKKLDRSTGTFPLAARNDVPLDSLQFIDFK-----RECL----SKGLHPRDLIG 72
Db      221 LYAW-DNRRTFRKSLGNDETDKKKGLGFFKVKNSKGCLLTTSPNSMHSRLTL 279
QY      73 SPITAFGKICTTS-----RLRLRLPGGEYEWQGINCRWRLLCAEVEKWCVCVHAT 125
Db      280 GPSLSLSGISGVSKSEMKRRAPPFGSGPPVQD-----KASE 318
QY      126 HLHGGSLWEILVCHSVRLKRRR-----PRPFFVGENSEDSSEEDHPA-FCVDP 174
Db      319 KVSLSGS-----QIDLQKKRAPADPPQPPPPSPFLIPNTEOKENKRSKMTWSLP 369

```

```

; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/316,856
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,185
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 1007
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte ID No: 3773014CD1
US-10-486-977-24

Query Match      3.8%; Score 113; DB 6; Length 1007;
Best Local Similarity 19.4%; Pred. No.1.8;
Matches 111; Conservative 61; Mismatches 171; Indels 228; Gaps 25;

Qy      146 KHRRRPR-----PFVGSNSDSSDHFAFCDFVTC-----TGASED----- 184
Db      4 RKQRPQQLISDCGPPASNGDASEDHQVCAKCAQFTDPTFEFLAHQNACTSDPPVM 63

Qy      185 ---SGDEGPPTRHSASGVQP-----VVDANADSPGSGDE-----GPSTRHSDSQ- 225
Db      64 VIIGQENPNNSASSPREPEGHNNPQVMDTEHSNPPDSGSSVPTDTMGPERGEESG 123

Qy      226 -----PPADETTVTDNVE----- 240
Db      124 HFLVAATGAAGGGGUILASPKLGAATPLPESTPAPPPPPPPPGVSGHNLPIIL 183

Qy      241 DDLTLIDK-----ESAC---ALMYHYGQ-----EMDMLMRACDEDLFDLLG 279
Db      184 EELRVLQRIHQWMTQEIQICRVLLGLSGIVGAPASPSLPGTCTASTKPLPLFS 243

Qy      280 IPEDV-----IATSPQGGDT-----DASGV----- 299
Db      244 PIKEVQTSKTLASSSSSSSSGAETPKQAFFHLYHPLGSHQPFPSAGGVGRSHKPTAPS 303

Qy      300 -----VTEGSIAASAVGA--GVEDVVLACALEAQNVAQVLEISD 338
Db      304 PALPGSTDLIASHLAPFTTGLLAQCLGARGLEATASGPLLKPQNSG---ELSY 359

Qy      339 EVDVDGAGLPPARRRRPVVGEFLWDGP-----RRH---ERP-----TTR---RIR 378
Db      360 GEVMGPLEKPGGRHKRCFCAPKAFGSDSALQIHLRSHTGERPYKCNVCGNFRFTTGNLKVH 419

Qy      379 HRKLRSAYYAVRPPVMTIDRLGVIFYFCRP-AMSL------VERKVFIL 423
Db      420 FHRREKIPVQNPVHPVPHLDVYITSSGLPYGMSVPPEKABEEAATPGGGVVERKELVA 479

Qy      424 CSQNPLADIHSHCLHSGK-----LRVLLPK-PDDNNNTGSG-DVNNLLAAV 466
Db      480 STTALSATESLITLLSTSAGTATAPGLPAFNKFLVLMKAVEPKNKADENTPGSGEISAIGV 539

Qy      467 LRFPAS-----GLVIVSLRSGIVYKVLCKST 492
Db      540 AESTATRMQLSKLVTSPLSPWALLTNHFKST 570

```

```

RESULT 6
US-60-556-841-7793
; Sequence 7793, Application US/60556841
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)
; CURRENT APPLICATION NUMBER: US/60/556,841
; CURRENT FILING DATE: 2004-03-25
; NUMBER OF SEQ ID NOS: 12463
; SEQ ID NO 7793
; LENGTH: 327

```

```
; TYPE: PRT
; ORGANISM: Oryza sativa
US-60-556-841-7793

Query Match
Best Local Similarity 3.7%; Score 112; DB 7; Length 327;
Matches 75; Conservative 42; Mismatches 132; Indels 82; Gaps 13;

QY 253 ALMYHVGQEMDMLRAMCDEDFLLIGIPEDVIATSPQGGDTDAGVVTGSGIAASAVGA 312
Db 53 ALAFHAGQ-----AVAIPIHDVIAPRKEGPI-AVGAQILAEAGAGLFS 95

QY 313 GVEDVILGALBAQNVAEYVLEISDEVDGAGLPPASRRPV-----VGEFLMD--- 363
Db 96 GVSATMLRQTLSTTRMGLYDI-LKKWTOENGGLPLRHKIAAGLIAAGVGAAGVGNPAD 154

QY 364 -----DGPRRHERPTTRIRHKLRSAYYRVARPFWMITDRLGVEVEFGRPAMSL 415
Db 155 VAMVRQADG-----RLPLAERNYRSVGDALGRMARDE-----GVSLWRG---SSLT 200

QY 416 VERKVFILCSQNLADIHSCLHSRKLRLVLLPKPDDNNTGPDVNLAAVLRFSAGLV 475
Db 201 VNRAMIVTASQATVDQAEKALARR-----GFGADGLGTHVVASFAAGLV 246

QY 476 IVSLRSGL-YVKNLCKSTVLYHGNPNPKKFGVIGLSSRAVLDFVNVAYRIQGEHIKK 534
Db 247 AAAANPDVNVKTRVNNMKVAPGAPPYSGAIDCAKLT-----VRSEG-----V 290

QY 535 TTVFTGSDFTSAEQ--FDWVPLVIXLRVSV 563
Db 291 MALYKGFPTVSRQGFPTWLPVTLQVRKV 321
```

```
RESULT 7
PCT-US04-07412-706
; Sequence 706, Application PC/TUS0407412
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Gezhi
; APPLICANT: Wang, Zhiwei
; APPLICANT: Boyle, Bryan J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 822CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/07412
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: US 10/389,559
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,264
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/340,187
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 10/296,115
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
```

```
; PRIOR APPLICATION NUMBER: US 10/275,027
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1920
; SOFTWARE: PCT-FL_genes Version 6.0
; SEQ ID NO 706
; LENGTH: 686
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-07412-706

Query Match
Best Local Similarity 3.6%; Score 106.5; DB 1; Length 686;
Matches 61; Conservative 17; Mismatches 86; Indels 71; Gaps 11;

QY 179 GAISEDSGDEG-----PSTRHSAGVQVPVDANADSPSGSDEG----- 216
Db 78 GAHGETEABEGAEVFGGEBTSGAQVEGA---SPGRGAQGEPRGEAQREPEDSAA 134

QY 217 PSTRHSDSQPPADETTVH-----TDNVEDDLTLDDKESACALMYHVQEMD 263
Db 135 PERQEABORPEVPEGSASGEAGDSVDAEGPLGDNIEAGPGAGDSVEAEG---RVGDSVD 191

QY 264 MLMRA-----MCDE-----DLFDLLIGIPEDVIATSPQGGDT-DASGVVTEGSI 305
Db 192 AEGPAGDSVDAEGPLGDNIAQEPAGDSVDAEGRVGSVDAGPAGDSVDAEGRVGSVE 251

QY 306 AASAVGAGVEDVYLAG-ALEAQNVAEYVLEISDEVDGAGLPPASRRRPVUGE 359
Db 252 AGDPAGDGEVAGVPAGDSVDAEAGPAGDSM-----DAEG--PAGRARRVSGE 295
```

```
RESULT 8
US-10-796-307-549
; Sequence 549, Application US/10796307
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001509
; CURRENT APPLICATION NUMBER: US/10/796,307
; CURRENT FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 44201
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 549
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-796-307-549

Query Match
Best Local Similarity 3.5%; Score 104.5; DB 6; Length 503;
Matches 111; Conservative 61; Mismatches 205; Indels 207; Gaps 26;

QY 11 EFIVGALDSKYPVKNWLDNRSTGTFLAP---AARNDVIPLDSLQFFIDFKRECLSKGLHP 67
Db 16 EMLLGEISSGCGYGLQWLDEATCTPRVPWKHFARKDLSEADAR---IPKAWAVARGRP 71

QY 68 RDLGSPITAFGKICTTSRRRLPGEYEVVQGINCRRLLCARVKECWCVCVHARTHL 127
Db 72 PSSRGGG-----PPPEAETA-----RAGKNTNFCAL 99

QY 128 HGSLSLWEILYQHSVRLKHEHRRPRPF-----VGENSDSSEEDHPAFCDVPV 175
Db 100 RS-----TRRFVMLRDNNSGDPADPHKVVYALSRELWCWREGTGTDTQEAAPAAVPPP- 150

QY 176 TGTGAESDSGDEGSTRHSAGVQVPVDANADSPSGSDEGSTRHSQPPADETTVH 235
Db 151 -----QGGPFGFLATHAGLQ-----AFG-----PLPAP----- 175
```

```

236 TDNVEDLTLDKESACALYHVGOEMDMLRAMCDEDLFOLLGIPEDVIATSQGGDT 295
176 -:::--ADKGDLLQAVQQSCLADHL-----LTASWGAD-- 204
296 ASGVVTEGSIAASAVGAVEVDVILAGALEAQNAGEYVLEISDEEVDGAGLP----- 348
205 -----PVPTKAPGSGOGLPLTGA-----CAG-----GPGLPAGELGYW 238
349 -----PASRRRP---VVGEFLWDGPPRHE---RPTTRIRHRKLRSAYYVARPPVMIT 397
239 AVETTPSGQPQAALTTCGAAPSPHQABPYLSFP-----SACTAVQEFS---P 286
398 DRUGVEVFYFCRPAMSLVERK---VFILCSONPLADISHSCLSHRSKGLRLVLLPKPDNNT 455
287 GALTVTMYKGRTVLQKVGHGSPCTFLYGPPDP-----AVRATDPOQVAFPSPAUL-- 337
456 GPGDVNL---LAAVURSFASGLUVISLRSGIVYKNLCKSTVLYHGNNPNPKKFG---VICGL 510
338 -PQOKQLRYTEELLRHVAPGLHLRGPLWARRMGMCKVYWEVGPPGSGASPSPTACIL 396
511 SSRAVLDVEN-----VAQYRIQGHEHIKKTTFVIG-GDPTSA 546
397 PRNCDTIFPRVFQELVEFRQRGRSPRYTYLGFQDLSA 440

RESULT 9
US-10-796-307-556
; Sequence 556, Application US/10796307
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CH001509
; CURRENT APPLICATION NUMBER: US/10/796,307
; CURRENT FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 44201
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 556
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-796-307-556

```

Query Match	3.5%;	Score 104.5;	DB 6;	Length 503;
Best Local Similarity	19.0%;	Pred. Nm. 3.2;		
Matches	111;	Conservative 61;	Mismatches 205;	Indels 207;
				Gaps 26;
Qy	11	EFIVGALDSKXYPLVKNWLDSTGTFAP--AARDNVIPDLSLOFFIDFKRECLSKGLHP	67	
Db	16	EWLUGEISSGCGYELQWLDARTCFRVWKHFARKDLSEADAR----IPKAWAVARGWP	71	
Qy	68	RDLGSPITAFGKICTTSRRLLRFLPGEYEVVQGINCRWRLLCAEVKECWCVHARTHL	127	
Db	72	PSSRGGG-----PPPEATAE-----RAGWKTNFRCAL	99	
Qy	128	HSSSSLWEILYQHSVRLEKHHRRRRPRPF-----VGENSDDSEEDHPAFCDVPV	175	
Db	100	RS-----TRREVMLRDNGSDPADPHKYALSRCLWREGPGPDQTEAEAPAAVPPP-	150	
Qy	176	TQTGAESDSDGEGPSTRHSASGVQVPVDDANADSPGSGDEGPGSTRHSQQPPADETTVH	235	
Db	151	-----QGGPPGPFLLATHAGLQ-----APG-----FLPAP-----	175	
Qy	236	TDNVEDDLTLDKESACALMTHVQGEWMDMLRAMCDEDLFLLGIPEDVIATSPGGDGT	295	
Db	176	-----AGDKGDLLLQAVQCSCLADHL-----LTASMGAD--	204	
Qy	296	ASGVVVTGSIITAASAVGAGVEDVYLAGALEAQNVAGEYVLEISDEEVDGAGLP-----	349	
Db	205	-----PVPTKAPGEGQGLPLTGA-----CAG-----GPGLPAGELYGM	238	
Qy	349	-----PASRRRP-----VVGEFLWDDGPRRHE---RPTTRIRHRKLRLSAYYRVARPPVMT	397	

```

239 AVETTPSPGPOAALTGEAAAPSPHQAESPYSLPSP-----SACTAVQEPS---P 288
398 DRLGVEVFYFQRPAMSLVERK--VFILCQNPLADISHSLRSRKLRLVLLFKPDNDNT 455
287 GALDVTIMYKGRITVLQKVGHPSCTFLYGGPDP-----AVRATDPQQVAFPSPAEL-- 337
456 GRGDVNL--LAAVLRSPASGLVIVLSRSGIYVKNLCKSTVLVHGNNPPKKFG---VICGL 510
338 -PDQQLRYTBELLRHVAPGLHLELRFPQJWARMGCKKYWEVGPGSGASPSPTACLL 396
511 SSRALVDVFN-----VAQYRIQGHGHIKTTVFIG-GDPTSA 546
397 PRNCDTPIFDPRVFFQELVEFRAFQRGRSPRYTILYFGQDLSA 440

RESULT 10
US-60-563-440-1193
; Sequence 1193, Application US/60563440
; GENERAL INFORMATION:
; APPLICANT: CHANG, Sheng-Yung
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/563,440
; CURRENT FILING DATE: 2004-04-20
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1193
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-563-440-1193

```

Query Match	3.5%; Score 104.5; DB 7; Length 503;
Best Local Similarity	19.0%; Pred. No. 3.2;
Matches 111; Conservative	61; Mismatches 205; Indels 207; Gaps 26
QY	11 EFIVGALDSKYVPLKWLDRSTGTFAP---AARNVDIPLDSLQFFIDFKRECLSKGLHP 67
Db	16 EWLLEISGCGEGLWLDENARCTFRVPWGHFARKDLEADAR---IFKANAVARGRP 71
QY	68 ROLLGSPITAFGKICTTSRRRLRPLGBEYVWOGINCRWRLLCAEVKWCWCVHARTHL 127
Db	72 PSSRGGG---PPPEAETAEE---RAGKTNFRCAL 99
QY	128 HGSLSLWEILYQHSVRLEKHHRRRRPFP---VGNSDSSBEDHPAFCDVPV 175
Db	100 RS-----TRRFVMLRDSGSDPADHKYVALSRCLWREGPGTDQTEAEAPAAVPPP- 150
QY	176 TOTGAESDSDGEGFSTRHSASGVQVDDANADSPGSDGEGFSTRHSQQPPADETTVH 235
Db	151 -----QGPPPGPFLATHAGLQ-----APG-----PLFAP----- 175
QY	236 TDNVEDDLTLDDKESACALMVHYGQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDT 295
Db	176 -----AGDKDILLQAVQOSCLADHL-----LTASWGAD--- 204
QY	296 ASGVVTEGSIASAASVAGVEDVYLAGALEAQNVAGYVLEISDEEVDGAGLP----- 348
Db	205 -----PVPTKAPGEQGLPLTGA-----CAG-----GFGLPAGELYGW 238
QY	349 -----PASRRSP---VVGFEFLWDGPPRHE---RPTRTRIRHKLSAYRVARPPVMIT 397
Db	239 AVETTPSGCPGPAALTTEGAAPSDHPQASPYLSPSP-----SACTAQEFS---P 286
QY	398 DBRLGEVYFYCFORPAMSLVERK--VFILCSQNPLADISHLSHRKGLRVLLPKPDNDNT 455
Db	287 GALDVTIMYKRTVLQKVGHSPCTFYLGPPDP-----AVRATDPOQVAFSPSAEL--- 337

QY 456 GPGDNL--LAALRSFASGLVIVSLRSGIYVKNLCKSTVLYHGNNPPKFG---VICGL 510
Db 338 -PDQQLAYTEELLRHVAPGLHLRGPQLWARRMGKCKVYVEVGGPGGASPTPACLL 396
QY 511 SSRVLDVN-----VAQRIQGHHEIKKTVIFIG-GDPTSA 546
Db 397 PRNCPTPIDFRVFFRELVEFRARQRSGSPRYTILYLGFGQDLA 440

RESULT 11

US-10-796-307-552
; Sequence 552, Application US/10796307
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001509
; CURRENT APPLICATION NUMBER: US/10/796,307
; CURRENT FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 44201
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 552
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-796-307-552

Query Match 3.5%; Score 104.5; DB 6; Length 516;
Best Local Similarity 19.0%; Pred. No. 3.3;
Matches 111; Conservative 61; Mismatches 205; Indels 207; Gaps 26;
QY 11 EFIVGALDSKYPVKWLDSTGTFLAP---AARNVDVPLDSLQFFIDFKRECLSKGLHP 67
Db 29 EWLLEISGCGYEGQLWLDDEARTCFRVWKHPARKDLEADAR----IFKAWAVARGRP 84
QY 68 RDLGSPITAFGKICTTSRRRLRPFGEYEVVQGINCRNRLLCAEVKECWCVCVHARTHL 127
Db 85 PSSRGGG-----PPPEAETA-----RAGWKTNFRCAL 112
QY 128 HGSLSLWELLYQHSVLEKHEHRRPRPF-----VGENSDSSEEDHRAFCVVPV 175
Db 113 RS-----TTRFVNLNRNSGDPADPHKVALSRELCWREGPGDQTEAEAPAAVPPP- 163
QY 176 TQTGAESDSGDEGSTRHSASGVQVDDANADSPGSDGEGPSTRHSQPPPADETTVH 235
Db 164 -----QGGPPGFLAHTAGLQ-----APG-----PLPAP----- 188
QY 236 TDNVEDDLTLDKESACALMTHVQEMDMLRAMCDEDLFDLLGIPEDVIATSPGGDTD 295
Db 189 -----AGDKGLLLQAVQQSCLDHL-----LTASWGAD-- 217
QY 296 ASGVVTEGSIASAAGVAGVEDVYLAGALEAQNVAGEYVLEISDEEVDGAGLP----- 348
Db 218 -----PVPTKAPGEQEGPLTGA-----CAG-----GPGLPAGELYGM 251
QY 349 -----PASRRP---VVGFEFLWDDGPRRHE---RPTTRIRHRKLRSAAYRVARPPVMIT 397
Db 252 AVETTPSPGPQPAALTGTGEAAAPESPQAEFVLSPP-----SACTAVQEPS---P 299
QY 398 DELGVEVEYFGRPAMSLVERK--VFILCSQNLADISHSCLHSRKGRLVLLPKPDNNNT 455
Db 300 GALDVTIMYKGTIVLQKVGHPSCTFLYGPDP-----AVRATDPOQVAFPSAEL-- 350
QY 456 GPGDNL--LAALRSFASGLVIVSLRSGIYVKNLCKSTVLYHGNNPPKFG---VICGL 510
Db 351 -PDQQLAYTEELLRHVAPGLHLRGPQLWARRMGKCKVYVEVGGPGGASPTPACLL 409
QY 511 SSRVLDVN-----VAQRIQGHHEIKKTVIFIG-GDPTSA 546
Db 410 PRNCPTPIDFRVFFRELVEFRARQRSGSPRYTILYLGFGQDLA 453

RESULT 12

US-10-784-004-389
; Sequence 389, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004

US-60-563-440-1190
; Sequence 1190, Application US/60563440
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: CHANG, Sheng-tung
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/563,440
; CURRENT FILING DATE: 2004-04-20
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1190
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-563-440-1190

Query Match 3.5%; Score 104.5; DB 7; Length 516;
Best Local Similarity 19.0%; Pred. No. 3.3;
Matches 111; Conservative 61; Mismatches 205; Indels 207; Gaps 26;
QY 11 EFIVGALDSKYPVKWLDSTGTFLAP---AARNVDVPLDSLQFFIDFKRECLSKGLHP 67
Db 29 EWLLEISGCGYEGQLWLDDEARTCFRVWKHPARKDLEADAR----IFKAWAVARGRP 84
QY 68 RDLGSPITAFGKICTTSRRRLRPFGEYEVVQGINCRNRLLCAEVKECWCVCVHARTHL 127
Db 85 PSSRGGG-----PPPEAETA-----RAGWKTNFRCAL 112
QY 128 HGSLSLWELLYQHSVLEKHEHRRPRPF-----VGENSDSSEEDHRAFCVVPV 175
Db 113 RS-----TTRFVNLNRNSGDPADPHKVALSRELCWREGPGDQTEAEAPAAVPPP- 163
QY 176 TQTGAESDSGDEGSTRHSASGVQVDDANADSPGSDGEGPSTRHSQPPPADETTVH 235
Db 164 -----QGGPPGFLAHTAGLQ-----APG-----PLPAP----- 188
QY 236 TDNVEDDLTLDKESACALMTHVQEMDMLRAMCDEDLFDLLGIPEDVIATSPGGDTD 295
Db 189 -----AGDKGLLLQAVQQSCLDHL-----LTASWGAD-- 217
QY 296 ASGVVTEGSIASAAGVAGVEDVYLAGALEAQNVAGEYVLEISDEEVDGAGLP----- 348
Db 218 -----PVPTKAPGEQEGPLTGA-----CAG-----GPGLPAGELYGM 251
QY 349 -----PASRRP---VVGFEFLWDDGPRRHE---RPTTRIRHRKLRSAAYRVARPPVMIT 397
Db 252 AVETTPSPGPQPAALTGTGEAAAPESPQAEFVLSPP-----SACTAVQEPS---P 299
QY 398 DELGVEVEYFGRPAMSLVERK--VFILCSQNLADISHSCLHSRKGRLVLLPKPDNNNT 455
Db 300 GALDVTIMYKGTIVLQKVGHPSCTFLYGPDP-----AVRATDPOQVAFPSAEL-- 350
QY 456 GPGDNL--LAALRSFASGLVIVSLRSGIYVKNLCKSTVLYHGNNPPKFG---VICGL 510
Db 351 -PDQQLAYTEELLRHVAPGLHLRGPQLWARRMGKCKVYVEVGGPGGASPTPACLL 409
QY 511 SSRVLDVN-----VAQRIQGHHEIKKTVIFIG-GDPTSA 546
Db 410 PRNCPTPIDFRVFFRELVEFRARQRSGSPRYTILYLGFGQDLA 453

RESULT 13

US-10-784-004-389
; Sequence 389, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004

; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 389
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-389

Query Match 3.5%; Score 104.5; DB 6; Length 1161;
Best Local Similarity 20.6%; Pred. No. 10;
Matches 100; Conservative 47; Mismatches 152; Indels 187; Gaps 26;

QY 157 GENSSEEDHPAFCDVPTQTGAESDSDGEGPSTRHSAGVQVDDANADSPGSGDEG 216
DB 135 GNATDEEEEEE-----EEGEEEEEEDDDDDDDSGAEIQDD---DEEGFDDDE 185
QY 217 --PSTRHSQPPADETTVHTDNVEDDLTL-----LDNEEELSELEERVEARKKTEKQSAALCVGVG--- 231
DB 186 EFDDDEHDDDD-----LDNEEELSELEERVEARKKTEKQSAALCVGVG--- 231
QY 264 MLMRAMCDEDLFDLLGI-----PE-----DVIATSQPGD---TDASGVVT 301
DB 232 ----SFADPD--DLPGLAHFLEHVMFGSLKYPDENGDFDAFLKKGSGDNASTDCERTVF 285
QY 302 EGSTAAASAVGAGVEDVLAGAL-----EAQNVAGEYVL-----EISD 338
DB 286 Q-----FDVQRKFKAALDRWAQFFHPLMIRDAIDREVAVDSEYQLARPSDANR 336
QY 339 EVDGAGLPPASRRP--VVGFLWDGPPRRHERPTTTRI--RHRKLRSAYYRVARPPVM 395
DB 337 KEMLFG-----SLARPGHMGKPFMGNAETLKHEPKKNNIDTHARLEFWMR----- 383
QY 396 ITDRLGVEVFFGPRPMSLEVERK-----VFILCSQNLADISHSLHSRKLRLVL 446
DB 384 -----YSAHYMTLVQSKETLDTLEKWTETFSQIP-----NNG----- 418
QY 447 LKPPDDNN--TGPQDVNLLAAVLR-----SFASGLVIV 477
DB 419 LKPNFSLHLPDPTPAFNKLYRVVPIRKIHATITWALPQQOQHYRVKPLHYISWLGVH 478
QY 478 SLRSGI--YVKNLCKSTVLYHGNPPKFGVICGLSSRAVLDFVNA--OYRIQGHHEHKK 534
DB 479 EGKGSILSYLRKCKWALALFGNGE-----TGFEQNSTYSVFSISITLTDGEYHEFYE 531
QY 535 T--TVF 538
DB 532 VAHTVF 537

RESULT 14

US-10-784-004-936
; Sequence 936, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201-6029-00000
; CURRENT APPLICATION NUMBER: US/10784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 936
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-936

Query Match 3.5%; Score 104.5; DB 6; Length 1161;
Best Local Similarity 20.6%; Pred. No. 10;
Matches 100; Conservative 47; Mismatches 152; Indels 187; Gaps 26;

QY 157 GENSSEEDHPAFCDVPTQTGAESDSDGEGPSTRHSAGVQVDDANADSPGSGDEG 216
DB 135 GNATDEEEEEE-----EEGEEEEEEDDDDDDDSGAEIQDD---DEEGFDDDE 185
QY 217 --PSTRHSQPPADETTVHTDNVEDDLTL-----LDNEEELSELEERVEARKKTEKQSAALCVGVG--- 231
DB 186 EFDDDEHDDDD-----LDNEEELSELEERVEARKKTEKQSAALCVGVG--- 231
QY 264 MLMRAMCDEDLFDLLGI-----PE-----DVIATSQPGD---TDASGVVT 301
DB 232 ----SFADPD--DLPGLAHFLEHVMFGSLKYPDENGDFDAFLKKGSGDNASTDCERTVF 285
QY 302 EGSTAAASAVGAGVEDVLAGAL-----EAQNVAGEYVL-----EISD 338
DB 286 Q-----FDVQRKFKAALDRWAQFFHPLMIRDAIDREVAVDSEYQLARPSDANR 336
QY 339 EVDGAGLPPASRRP--VVGFLWDGPPRRHERPTTTRI--RHRKLRSAYYRVARPPVM 395
DB 337 KEMLFG-----SLARPGHMGKPFMGNAETLKHEPKKNNIDTHARLEFWMR----- 383
QY 396 ITDRLGVEVFFGPRPMSLEVERK-----VFILCSQNLADISHSLHSRKLRLVL 446
DB 384 -----YSAHYMTLVQSKETLDTLEKWTETFSQIP-----NNG----- 418
QY 447 LKPPDDNN--TGPQDVNLLAAVLR-----SFASGLVIV 477
DB 419 LKPNFSLHLPDPTPAFNKLYRVVPIRKIHATITWALPQQOQHYRVKPLHYISWLGVH 478
QY 478 SLRSGI--YVKNLCKSTVLYHGNPPKFGVICGLSSRAVLDFVNA--OYRIQGHHEHKK 534
DB 479 EGKGSILSYLRKCKWALALFGNGE-----TGFEQNSTYSVFSISITLTDGEYHEFYE 531
QY 535 T--TVF 538
DB 532 VAHTVF 537

DB 135 GNATDEEEEEE-----EEGEEEEEEDDDDDDDSGAEIQDD---DEEGFDDDE 185
QY 217 --PSTRHSQPPADETTVHTDNVEDDLTL-----LDNEEELSELEERVEARKKTEKQSAALCVGVG--- 231
DB 186 EFDDDEHDDDD-----LDNEEELSELEERVEARKKTEKQSAALCVGVG--- 231
QY 264 MLMRAMCDEDLFDLLGI-----PE-----DVIATSQPGD---TDASGVVT 301
DB 232 ----SFADPD--DLPGLAHFLEHVMFGSLKYPDENGDFDAFLKKGSGDNASTDCERTVF 285
QY 302 EGSTAAASAVGAGVEDVLAGAL-----EAQNVAGEYVL-----EISD 338
DB 286 Q-----FDVQRKFKAALDRWAQFFHPLMIRDAIDREVAVDSEYQLARPSDANR 336
QY 339 EVDGAGLPPASRRP--VVGFLWDGPPRRHERPTTTRI--RHRKLRSAYYRVARPPVM 395
DB 337 KEMLFG-----SLARPGHMGKPFMGNAETLKHEPKKNNIDTHARLEFWMR----- 383
QY 396 ITDRLGVEVFFGPRPMSLEVERK-----VFILCSQNLADISHSLHSRKLRLVL 446
DB 384 -----YSAHYMTLVQSKETLDTLEKWTETFSQIP-----NNG----- 418
QY 447 LKPPDDNN--TGPQDVNLLAAVLR-----SFASGLVIV 477
DB 419 LKPNFSLHLPDPTPAFNKLYRVVPIRKIHATITWALPQQOQHYRVKPLHYISWLGVH 478
QY 478 SLRSGI--YVKNLCKSTVLYHGNPPKFGVICGLSSRAVLDFVNA--OYRIQGHHEHKK 534
DB 479 EGKGSILSYLRKCKWALALFGNGE-----TGFEQNSTYSVFSISITLTDGEYHEFYE 531
QY 535 T--TVF 538
DB 532 VAHTVF 537

RESULT 15

US-60-565-632-9051
; Sequence 9051, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Muryikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)B
; CURRENT APPLICATION NUMBER: US/60/565,632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 9051
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
US-60-565-632-9051

Query Match 3.5%; Score 104; DB 7; Length 213;
Best Local Similarity 25.4%; Pred. No. 1.1;
Matches 47; Conservative 11; Mismatches 79; Indels 48; Gaps 7;

QY 163 SEEDHPAFCDVPTQTGAESDSDGEGPSTRHSAGVQVDDANADSPGSGDEG 210
DB 23 SDPDPS-----PDEQTNAPSVSTSDSPDVTDSSTAPDTSDTDPDPTDSSSTAPTOSPP 78
QY 211 GSGEGPSTRHSQPPFADP--TTVHTDNVEDDLT-----LLDKESACAL 254
DB 79 NPATDSSSTAPTDSPPDPTDSSSTAPTDSSTAPTDSSTAPTDSSTAPTDSSTAPTDS 135

QY 255 MYHVQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTDASG-----VVTEGSIA 306
Db 136 -----SPTDSPPEPATDSSASPTDSPDSTTASSGSSSTLSSDSTGATPESTKGSAA 190
QY 307 ASAVG 311
Db 191 SLFTG 195

Search completed: June 9, 2004, 09:07:37
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 08:58:25 ; Search time 21 Seconds
(without alignments) 2597.171 Million cell

Title: US-10-656-499-2

Perfect score: 2993
Sequence: 1 MAGRRLTWISFIVGALDSD.....QFDMVPLVIKRLRSVTCDD 567

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : PIR 78: *

```
1: ̄pir1: *
2: ̄pir2: *
3: ̄pir3: *
4: ̄pir4: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	121.5	4.1	1166	2	T15628	hypothetical prote
2	117.5	3.9	1567	2	T03730	antigen containing
3	117	3.9	1316	2	T00381	KIAA0633 protein -
4	115.5	3.9	3436	2	S58559	tegument protein 6
5	113.5	3.8	1005	2	C71513	hypothetical prote
6	111	3.7	1216	2	A58620	apical endosomal p
7	109	3.6	1664	2	F84485	probable retroelem
8	109	3.6	4717	2	T41581	hypothetical coile
9	108	3.6	953	2	T70681	probable rne prote
10	107	3.6	836	2	B84417	cell surface glyco
11	107	3.6	883	2	S57653	breivican precursor
12	107	3.6	974	2	A40580	lodestar maternal-
13	106.5	3.6	347	2	T35013	probable membrane
14	106.5	3.6	1196	2	S35994	DNA repair protein
15	105.5	3.5	827	2	T04789	hypothetical prote
16	105	3.5	813	2	A13651	two component sens
17	105	3.5	1275	2	H97433	non-motile and pha
18	105	3.5	1892	2	T18314	hypothetical prote
19	104.5	3.5	1161	2	I59311	nardillyasin (EC 3.4
20	104	3.5	852	2	A28459	cell surface glyco
21	103.5	3.5	273	1	EDBE12	immediate-early pr
22	103.5	3.5	284	2	T43608	immediate-early pr
23	103	3.4	357	2	P70898	probable gcp-bindin
24	103	3.4	350	2	S24773	URSL1 protein - sm
25	102.5	3.4	984	2	S84781	hypothetical prote
26	102.5	3.4	1506	2	A96808	hypothetical prote
27	102.5	3.4	2124	2	A28452	proteoglycan core
28	102	3.4	690	2	H84309	protoporphyrin ix
29	102	3.4	925	2	P75222	protein P7522.2 li
30	102	3.4	9494	2	H96494	

30	102	3.4	1085	2	S55352	IFH1 protein - yeast
31	101.5	3.4	863	2	S06017	neuraxin - rat
32	101.5	3.4	897	2	T02808	conserved hypothet
33	101	3.4	1092	2	T18305	replication factor
34	101	3.4	1092	2	T18305	replication factor
35	101	3.4	3415	2	A46105	polyprotein(NS1, N
36	100.5	3.4	2364	2	A56577	microtubule-associ
37	100	3.3	425	2	A45064	interferon consens
38	100	3.3	465	2	H84280	hypothetical prote
39	99.5	3.3	593	2	S49525	glycoprotein G - s
40	99.5	3.3	1095	2	S76044	hypothetical prote
41	99.5	3.3	1558	2	T29253	hypothetical prote
42	99.5	3.3	1998	2	T13009	hypothetical prote
43	99.5	3.3	2327	2	T42630	aggrscan - bovine
44	99	3.3	1058	2	T30178	mitotic checkpoint
45	99	3.3	1102	2	T31004	probable mitotic c

ALIGNMENTS

RESULT 1

T15628
hypothetical protein C25H3.9 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T15628
R:Johnson, D.
submitted to the EMBL Data Library, June 1995
A/Description: The sequence of *C. elegans* cosmid C25H3.
A/Reference number: Z18379
A/Accession: T15628
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1166 <JUH>
A/Cross-references: EMBL:U29835; NID:G868251; PID:G868258; PIDN:AAA68787.1; CESP:C25H3.9
A/Experimental source: strain Bristol N2
C/Genetics:
A/Gene: CESP:C25H3.9
A/Intons: 45/3; 82/3; 129/2; 180/2; 202/1; 215/3; 252/1; 374/3; 544/2; 1028/3; 1108/1

	Query Match	4.18; Score 121.5; DB 2; Length 1166;
	Best Local Similarity	20.0%; Pred. No. 0.83;
	Matches 107; Conservative	83; Mismatches 193; Indels 152; Gaps 25;
QY	69 DLIGSPITAFGKICTTSRRRLRLLPGGEYEYVGGINCRWRLLCA--SVKECWMCVHAR- 124 : :: :	: :: :
Db	502 DILGLEUFRH-----QMTQURAVNRWARFERRAQHRKRWRPLCTTGNAKEWKFAYNRYV 556 : :: :	: :: :
QY	125 ---TLHSGGSLMEILLYQHVSRLLEKKRRPRPFVGENSDSSEEDHPAFCDPVPVTQTGAE 181 : :: :	: :: :
Db	557 LEESRRANANRTWDFAHERRASRNAYRAYRKLIIG-----LIANPNAIQ 601 : :: :	: :: :
QY	182 SEDSGDEGPSTRHASGVQVDDANADSGSGDGEPSRTHSDSQPPADETIVHTDNVED 241 : :: :	: :: :
Db	602 STVSTDSSNTFPGNALAWPVTTSN-----GASSNHNQSL-----ESTAIKKQIER 647 : :: :	: :: :
QY	242 D-----LTLLDKESACALMYHVGOEMLMRAMCEDLFDLLGIPEDVIATISQPGDIT 294 : :: :	: :: :
Db	648 DAQTYTHELHLFRETVPKRIREKEXELGITAAPIDTEDAFETLFPPDELIIVDEPTAPT 707 : :: :	: :: :
QY	295 DASGVVTGEGTAASAVG-----AGVEDVYLAGALEQNRY--AGEYVLE 335 : :: :	: :: :
Db	708 EP----TNGLGYCHMITFFGQAQQDEKQEDNKFDGNVDV--GELKDINKEMEIEELD 760 : :: :	: :: :
QY	336 ISDSEVDGDAGLPASRRRRPVVGEFLWDGDRRHERPTTRIRHKLR-----SAFYRVAR 391 : :: :	: :: :
Db	761 VLRHSWDDSTLL-----RRDALIAQI-----SLRLEHTLRFVDSVDHDGIEQ 803 : :: :	: :: :
QY	392 PPMVIMTRLRGVEVFYGRPA---MSLEV---ERKVFIICSQNP-----LADISHLSLHS- 439 : :: :	: :: :
Db	804 ORVLALIELSGYSRRWELSPKHLYSDVTDVMWSVORLRSRSHPRKPSFAELSLSLAYST 863 : :: :	: :: :

Query Match 3.9%; Score 115.5; DB 2; Length 3436;
Best Local Similarity 21.4%; Pred. No. 9.8;
Matches 79; Conservative 38; Mismatches 132; Indels 121; Gaps 18;

QY 148 RRRPRRFVGENSDSEEDHPAFCDVPTGTGAESDSDGEGPSTRHSASG-----VQ 200
Db 410 RTAPETVWGDGNSFDFH-----LPTTRD--PGRCGRVTAADHAVGASLGVESLQ 462
QY 201 P-----VDDANADSPGSDGEGPSTRHSQSPPADETVHTDVEDD----- 242
Db 463 PPTTLADERSGRSGDGRGSPSA-----ADDRVGVNNSDNGPPTTPPADDES 513
QY 243 --LTLDKESACALMYHVGQEMDMRAMCEDLFDLGLIPEDVIATSPQGGTDSAG-V 299
Db 514 RRLSLRLEGEG-----DAGTASEDV--AVFGDSTRRAGEK 546
QY 300 VTEGSIASAASVGA-----GVRDVIAG-----ALEAQNVAAGEVYLEISDEVD 343
Db 547 VTHPTGSPARGSRSPGVPAED-----AGKAGSPGHRVPSARSQAQVGSFNSQAG 603
QY 344 GAGLPP-----ASRRPVVGE-----FLWDDGPRRH-ERTTTRIRHRKLSAYYV 389
Db 604 DEGLPRDTRGKKAARRSPGGGAPAFNLHGDVNSHKEPPRPE-----KT 654
QY 390 ARPVMITDRLGVEFFVFGPAMSLVERKVFILCONPLADISHCLSRKGLRVLKP 449
Db 655 GEDPSMVKTAGDMSPLYGDPSTGLMHK-----CSDRPTFESIHACTRELDEMVL 706
QY 450 PDNNITPGD 459
Db 707 DEDGGSGCD 716

RESULT 5
C71513
hypothetical protein CT456 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: C71513
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: C71513
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1005 <ARN>
A;Cross-references: GB:AE001319; GB:AE001273; NID:G3328881; PIDN:AAC68056.1; PID:G332888
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: CT456

Query Match 3.8%; Score 113.5; DB 2; Length 1005;
Best Local Similarity 21.0%; Pred. No. 2.7;
Matches 58; Conservative 41; Mismatches 88; Indels 89; Gaps 13;

QY 176 TGTGAESDSDG-----DEGPSTRHSASGVQ-PVDDANADSPGS-----GDEG----- 216
Db 491 TQAGPSSDDGISFNETPCAGPAAPSPPTPSIPILNVNVNNGTNVIGTNVNTNT 550
QY 217 -PSTRHSQSPPADETVHTDVEDDLTLDKESACA-----LMYHVGQEMDM 266
Db 551 TPTTQSTDASTDSDIDINTNNQTDINTDKSDGAGVNGDISSETSSGSDSGSVS 610
QY 267 RAMCED-----LFDLLGTP-----DVATSPQGGTDA-----SGVTE 302
Db 611 SSSDDKNASVNGDGPAMKOILSAVRKHLDVYPGNGSGTEGFLPANQTLGDVSDVENK 670
QY 303 GS-----IAASAVGAGVEDVYLAGALEAQNVAAGEVYLEISDEEVDGGA 345
Db 671 GSAQDTKLSGNTGAGDDDTTAAVNGAEEITLS-----DTSIGIDGDSVDSRSSGDESG 727
QY 346 GL---PPASRRRPVVEFLWDDGP-----RRH 369

RESULT 7
P84485
probable retroelement pol polyprotein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: P84485
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

Db 728 GVSPSSSSNKNVTAVG-----NDGPSGLDILAAVRKH 759

RESULT 6

A55620
apical endosomal protein precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 05-Nov-1999
C;Accession: A55620
R;Speelman, B.A.; Allen, K.; Grounds, T.L.; Neutra, M.R.; Kirchhausen, T.; Wilson, J.M.
J. Biol. Chem. 270, 1583-1588, 1995
A;Title: Molecular characterization of an apical early endosomal glycoprotein from devel
A;Reference number: A55620; MUID:95130529; PMID:7823488
A;Accession: A55620
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1216 <SPE>
A;Cross-references: GB:L37380; NID:G642652; PIDN:AAA65200.1; PID:G777776
C;Superfamily: LDL receptor ligand-binding repeat homology; MAM homology
F:231-267/Domain: LDL receptor ligand-binding repeat homology <LDLA>
F:456-489/Domain: LDL receptor ligand-binding repeat homology <LDLB>
F:654-813/Domain: MAM homology <MAM1>
F:819-973/Domain: MAM homology <MAM2>

Query Match 3.7%; Score 111; DB 2; Length 1216;
Best Local Similarity 20.4%; Pred. No. 5.4;
Matches 99; Conservative 49; Mismatches 154; Indels 184; Gaps 25;

QY 30 RSTGTFAPAARNVDVPLDSLOFFIDFKRECLSKGLHPRDLLGSPITAFGKICTSTRRLR 89
Db 708 RQGGALLTRPQVFPV-----KECLSFYH---LHGPI---GTLCLAMR-- 748
QY 90 RLPGEYEVV--QGNCRRRLCAEVKECWCWCVHARHLHSGSSLWEIL-----YQ 139
Db 749 --EGEDTLNRSRGTGNRH-----QAWTLH---HOLQSTKYQLLEGLRQYH 796
QY 140 HSVLEKHRRPRRPFVGENSDSEEDHPAFCDVPTGTGAESDSDGEGPSTR-HSASG 198
Db 797 GTMGDDMAVRFGPCAAKRCSPEDSD---CGF-----SPDGLWLTQNNASG 842
QY 199 VQP-----VDDANADSPGS---GDSP-----STHSQSOPPAPDETTHVNVEDDL 243
Db 843 LGPWGFWIDHTTGTGAQGHYVVDVTSNLLPKGVASLTSEHP----- 886
QY 244 TLDKESACALMYHVGQEMDMRAMCEDLFDLGLIPEDVIATSPQGGTDSAGVVT-- 301
Db 887 --LSRPACLSFWYKLSFNPCTLVFVEE-----STRQELSSGHGGFAWRLGSVNQ 938
QY 302 ---EGSIAASAVGAGVEDVYLAGALEAQNVAAGEVYLEISDEEVDGAGLPPAS----- 351
Db 939 AEQAMKVVFEMASGVESYNA-----LDDISLQDGPACQPGSCDFESGL 983
QY 352 ---RRRPVVGFLWD---DGPRRHERPTTRIRHRKLSAYYVARPVMITDLGV 402
Db 984 CGWHLFWPLGGVSWSSGATSPRPS-----VDHTVGT 1021
QY 403 EVFYF-----GRPAMSLEVERKVFILCONPLADISHSL-----HSRKG 442
Db 1022 EAGHFAPFETSVLPGCGAAWLGSE-----PLPATAVSCLHFWYMGPPAFYKG 1071
QY 443 -LRVLL 447
Db 1072 ELRVLL 1077

RESULT 7

P84485
probable retroelement pol polyprotein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: P84485
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84485
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1664 <S>
A:Cross-references: GB:AE002093; NID:94309765; PIDN:AA15534.1; GSPDB:GN00139
C:Genetics:
A:Map position: 2

Query Match 3.6%; Score 109; DB 2; Length 1664;
Best Local Similarity 21.4%; Pred. No. 12;
Matches 126; Conservative 61; Mismatches 215; Indels 186; Gaps 31;
QY 78 FGKICTTSSR-----RLRRPGEEYEV---VQGINCRWRLLCAEVKECWCVHAR 124
DB 166 FGLLLTSTYNDLLKHLRADKLPNLB-EVCSQIQEQANGYKYNKNAKALWCBECK 224
QY 125 THLHSGSLWEILYOHSLVLEKRRPR-RPFVGENSSBEDHPAFCDVPTGTGAESE 183
DB 225 RSGHTKECW-TLHPH---LRPGREPRANVTGENFTQSE-----QSGTSSNQ 268
QY 184 DSGDGPSTRHSASGVQPDV-----DANADSPGS-----DEGPSTEH--SDS 224
DB 269 HLGNGAMAMASSDILVRSDLKALIKESGKSYHALSKPLIIDSGAS-HHMTSDS 327
QY 225 -----QPPADETTHTDNVE-----DDLTLDKESACALMYHVGQEMDMRAMCDEDL 274
DB 328 KLIISIEPALGNVVIANGDRIPVKGVGLDLFDKSSKAFYPTFTSNLLSVKATTDLNC 387
QY 275 FDLGIGPEDV---IATQPGDDTASGVVTEGSTAASVAGVEDVYLAGALEAQNAG 330
DB 388 YAIFG-PNEVHFQIETSR-----VLGGV-----TKOG 415
QY 331 EYVLISDEEYDAGLPPASRRPVPVGEFLWDGPRRHERPTTIRIRHKLRSAYRVA 390
DB 416 LVVLE-----DTPSVPLSHFSILGNANSESHALGHPHSRAK-LLLPSTSEKND 468
QY 391 RPPVMITRLGVVEYFGRPANSLEVERKVFILCSQNPLADISHS-----CLHSRKGRL 444
DB 469 ECEACILGHCKSVF-----PKSSTIYEKCF-----DLHSDVMTSPCL-SRENHK 513
QY 445 -----VLLPKPDNNTGSDVNLAAVLRSFASGLVTS-----LR 480
DB 514 YVTFIDSKSTFTLLPSKD-----RVLEAFTNQTVTYTHYDAKILR 560
QY 481 S---GIYVKNLCKSTVLVHG-----NNPPKFGVIGLSSRAVLDPVFNVAQVRIQGE 530
DB 561 SDNRGEYTSFAFKOHLNKHGIIHQTSCTPYTPQONGV-AERKNRHLMEYRVMFHTNVPK 619
QY 531 H-----IKKTVFTGGPTSAEQFDMV-PLVILKLSRVTC 565
DB 620 HFWDGVVSACYLINTPTKILLSSPFEVLNKPFPINHLRVFGVCV 667

RESULT 8
T41581
hypoetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-Aug-2003
C:Accession: T41581
R:Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z22002
A:Accession: T41581
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4717 <M>
A:Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08

A:Experimental source: strain 972h(-)
C:Genetics:
A:Gene: SPDB:SPCC737.08
A:Map position: 3
C:Superfamily: midasin (AAA ATPase with von Willebrand factor type A (vWA) domain)

Query Match 3.6%; Score 109; DB 2; Length 4717;
Best Local Similarity 22.0%; Pred. No. 46;
Matches 72; Conservative 26; Mismatches 110; Indels 120; Gaps 13;

QY 158 ENSDSSEED-----HPAFCDVPTGTGAESDSDGSGPSTRHSASGV 199
DB 3973 ENSDSEENQDLDEEVNDIPEDLSNLSNKLWDENEEDLLTEQKSNEQSANNESDLV 4032
QY 200 QPVDDANA-----DSPGSGDE-GPSTRHSDSQPPPADTTVHTDNVED- 241
DB 4033 SKEDDNKALDKDROKDEDEEEMSDVGIDDBITQDPIQENNSQPPPENED--HLDLPEDL 4090
QY 242 -----DLTLDKESACALMYHVGQEMDMRAMCD-----EDL 274
DB 4091 KLDEKGDVSKSDLEDMDEAENKKEADAEKQEPQDFPLEENNTLDEDTQQDDF 4150
QY 275 FDL-----GIPEDVIATSQPGD-----TDA 296
DB 4151 SDLAEDDEKMNEDGPFENVOENEESTEDGVKSDEELEGVEVPEQDAIDNHPKMDAKSTFA 4210
QY 297 SGVVTGSGTAASAVG-----AGVEDVYLAGALEAQNVA-GEYVLETSDEEVDG----- 344
DB 4211 SAADSENTDKGIVGENEELGEBDGAESGVGRNGTAGDEF---SSAEQVQKGETSTPTK 4267
QY 345 AGLPPASRRRPVVGEPFL--WDDGPRRHE 370
DB 4268 EAMSEADROYQSILGDLHREMQANRIHE 4295

RESULT 9
B70681

probable rne protein - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70681
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70681
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-953 <COL>
A:Cross-references: GB:Z81451; GB:AL123456; NID:93261662; PIDN:CAB03775.1; PID:e280523; I

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: rne

Query Match 3.6%; Score 108; DB 2; Length 953;
Best Local Similarity 22.0%; Pred. No. 6.6;
Matches 106; Conservative 50; Mismatches 166; Indels 160; Gaps 24;

QY 83 TTSRR-----LRRLPG---EEYEVQGINCRWRLLCAEVKECWCVHARTHLHSGSLWE 135
DB 34 TTSRRVLDALTALDGRVRSASHTVDRVDAVRVDLLA-----THLETAGVL-- 79
QY 136 ILYQHSVRLEKRRPRRPPFVGENSDSSEED-----HPAF-CDVPVPTGTGAESDSDGDE 188
DB 80 ---AASVHAPEASEBPESRLMLETQETRNADVERHYMFLFVAPQPIPEPLADDEDV-DD 135
QY 189 GPSTRHSASGVQPDVNDANADSPGSGDEGPSTH-----SSQP----- 226
DB 136 GPD-----YVADSDADDEGQLDRPANRRRREGRGGRGEGQSGDGFVDQOSE 187

QY 227 PPA-----DETVHTDNVEDDLTLLD-----KESACALMYH 257
Db 188 PRAQFTSADAEFDGDDRRDSEDEAGDNGEDENGSLAGNRERRRRKAS-----242
QY 258 VQEWDMMLMRANCDDELFDLLGIPEDVIATSPQGGDTASGVVTEGSIASAVGAGVEDV 317
Db 243 -GDNDAALEGFLDD-----PENTVHERVPRAGDKAGNSQDGG-----281
QY 318 YLAGALRAQNVAGEVYLEISDEEVDGAGLPPASRRPVV---GEFLWDDGPRRH---ERP 372
Db 282 --SSTIKIGDGTRELAQRDRD---DAGRRPPVLSEAEFL---ARREAVRV 332
QY 373 TTRIRHRKLSAYRVARPV-----MITDLRGVEVFPGRPAMSLVERKVFIIL 423
Db 333 MVRDRVR-----TEPPLGTRVYQIAVLED--GIVVEHFTVTSAAASLVGNVILG 381
QY 424 CSQNLADISHSLHSRKLRLVLPKPDNNTPGVDVNLAAVLRSFASGLVIVLSRGI 483
Db 382 IVQNVLPSEAFVDFIGRNGVL-----YAGEVNDADAGLGG-ADRKIEQALKPGD 432
QY 484 YV 485
Db 433 YV 434

RESULT 10
B84417
cell surface glycoprotein [imported] - Halobacterium sp. NRC-1
C/Species: Halobacterium sp. NRC-1
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C/Accession: B84417
R;NG, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Oner, A.D.; Ehardt, H.; Lowe, T.M.; Li
A/Title: Genome sequence of Halobacterium species NRC-1.
A/Reference number: B84160; MUID:20504483; PMID:11016950
A/Accession: B84417
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-836 <STO>
A/Cross-references: GB:AE004437; NID:g10882055; PIDN:AAG20702.1; GSPDB:GN00138
C/Genetics:
A/Gene: csf
C/Superfamily: S-layer glycoprotein

Query Match 3.6%; Score 107; DB 2; Length 836;
Best Local Similarity 24.0%; Pred. No. 6.5;
Matches 56; Conservative 30; Mismatches 67; Indels 80; Gaps 13;

QY 182 SEDSGDEGSPTRHSGSV-----SVASGATIFQGBEDVTFRKLNDNEKVPATLSRTGSGDEGVPLQMP 216
Db 42 SEDGKTEG-----SVASGATIFQGBEDVTFRKLNDNEKVPATLSRTGSGDEGVPLQMP 97

QY 217 --PSTRHSDSOPPADE---TTVHT-----DNVEDLT-----LLDKESACAL-- 254
Db 98 EDQSTGVSVDNGPNDDEADFGVTVQSPSVTLMVRNADNDVTGVLNTQODESSIAVDY 157

QY 255 MYHVQEMDMMLMRANCDDELFDLLGIPEDVIATSPQGG-----DTASG 298
Db 158 NYVAAEDLELAVE-----DEGLD---VTDEILAAQSGGAYEDGTGNGNPNTRFPDIDPN 211

QY 299 VYTEGSIASAVGAGVEDVYLAGALEARQNV---AGEVYLEISDEEVDGAGL 347
Db 212 -VDAGDYTVSV--EGVEDLDFGDATESASVTISSNKAISLAEDVEVVOGANTL 261

C/Accession: S57653
R;Rauch, U.; Forsberg, N.; Kulbe, G.; Arnold-Ammer, I.; Faessler, R.
submitted to the EMBL Data Library, May 1995
A/Description: Cloning and sequence of mouse neurocan and brevican and their different e
A/Reference number: S57653
A/Accession: S57653
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-883 <RAU>
A/Cross-references: EMBL:X87096; NID:g886889; PIDN:CAA60575.1; PID:g886890
C/Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
F;49-138/Domain: immunoglobulin homology <IMM>
F;173-250/Domain: link protein repeat homology <LNK1>
F;271-352/Domain: link protein repeat homology <LNK2>
F;626-657/Domain: EGF homology <EGF>
F;664-784/Domain: C-type lectin homology <LCH>
F;791-847/Domain: complement factor H repeat homology <FHD>

Query Match 3.6%; Score 107; DB 2; Length 883;
Best Local Similarity 24.0%; Pred. No. 7;
Matches 78; Conservative 19; Mismatches 104; Indels 124; Gaps 16;

QY 151 PRPFVGEN-----SDSSEEDHPA-----PCDVPVTQTGAESDSDG-----DE 188
Db 423 PRPFLESETQSIAPPTSESEEGVALESEERFKDLEAEEKEQEDLWVWPRELSSPLPT 482

QY 189 GPSTRHSGVQVPVDANADSPGSGDEGSTRHSDSQPPPADETVHTDNVEDDLTLLDK 248
Db 483 GSETEHSLSQVSPPAQVQLDASPSGPPR-----FRGPPAE-----TLLP- 524

QY 249 ESACALMYHVQEMDMMLMRANCDDELFDLLGIPEDVIATSPQGGDTASGVVTEGSIAS 308
Db 525 -----PREWSATSTEGG-----AR 538

QY 309 AVGAGVEDVYLAGALEARQNVAGEVYLEISDEEVDGAGLPPASRRRPVVGFLWDD-GPR 367
Db 539 EVGGETGSPSLGVPRESEAGSSILE-----DGPSLLPAT-----WAPVGPR 581

QY 368 RHERPTTRIRHRKLSAYRVARPVMTDLR---GVEVYFGRPAMSLVERKVFIILC 424
Db 582 ELETSPSEKSGRTVL--AGTSVQAQPVLPDTSASHGVAV-----APSSG-----DC 626

QY 425 SQNLADISHSLHSRKLRLVLP 448
Db 627 IPSCHN-OGTCLSEKSGPRCLCLP 650

RESULT 12
A40580
lodestar maternal-effect protein - fruit fly (Drosophila melanogaster)
N/Alternate names: probable nucleoside triphosphate binding protein lodestar
C/Species: Drosophila melanogaster
C/Date: 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 21-Jul-2000
C/Accession: A40580; MUID:92009170; PMID:1916263
R;Girdham, C.H.; Glover, D.M.
Genes Dev. 5, 1786-1799, 1991
A/Title: Chromosome tangling and breakage at anaphase result from mutations in lodestar,
A/Reference number: A40580; MUID:92009170; PMID:1916263
A/Accession: A40580
A/Molecule type: mRNA
A/Residues: 1-974 <GIR>
A/Cross-references: GB:X62629; NID:g8187; PIDN:CAA44496.1; PID:g8188
C/Genetics:
A/Gene: lodestar
A/Cross-references: FlyBase:FBgn0002542

Query Match 3.6%; Score 107; DB 2; Length 974;
Best Local Similarity 22.7%; Pred. No. 8;
Matches 82; Conservative 52; Mismatches 149; Indels 78; Gaps 17;

QY 158 ENSDSSEED--HPAFCDVPVTQTG--AESEDSGDE-----GPSTRHSGASGVQPV---D 203
Db 82 QSESEDSSEEDSVRPSARNTKRLPLGIPSESEDEDELEQALSPSTRMSITGVYRPQDLSD 141

Search completed: June 9, 2004, 09:03:11
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 08:51:34 ; Search time 18 Seconds
(without alignments)
1640.210 Million cell updates/sec

Title: US-10-656-499-2

Perfect score: 2993

Sequence: 1 MAGRLTWISFIVGALDSD.....QFDNVPVLKRLRSVTCD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141691

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121.5	4.1	2517	1 NCR2 HUMAN	Q9Y618 h nuclear r
2	118	3.9	1004	1 PTPX RAT	Q63475 rattus nov
3	116.5	3.9	466	1 X10 MOUSE	Q04888 mus musc
4	113.5	3.8	1005	1 Y456 CHLTR	Q84462 chlamydia t
5	113	3.8	1007	1 SAL2 HUMAN	Q9Y467 homo sapien
6	112.5	3.8	466	1 SAL2 HUMAN	Q56693 homo sapien
7	112.5	3.8	466	1 X10 RAT	Q88b91 mus musc
8	111	3.7	1161	1 NRDC MOUSE	Q63191 rattus nov
9	111	3.7	1216	1 AEGP RAT	Q9X96 mus musc
10	109.5	3.7	1004	1 SAL2 MOUSE	Q92985 homo sapien
11	108.5	3.6	503	1 IRP7 HUMAN	Q61361 mus musc
12	107	3.6	883	1 PGBB MOUSE	Q894h3 bifidobacte
13	107	3.6	969	1 PTKB_BIFLO	Q8C5W0 mus musc
14	107	3.6	1052	1 CLMA_MOUSE	Q34739 drosophila
15	107	3.6	1061	1 LDS_DROME	Q96n77 homo sapien
16	106.5	3.6	704	1 CLIE HUMAN	Q14629 xenopus lae
17	106.5	3.6	1196	1 XPG_XENLA	Q28343 canis fami
18	106.5	3.6	2333	1 PGCA CANFA	Q9TJ02 rattus nov
19	105.5	3.5	738	1 S521 RAT	Q47245 rattus nov
20	104.5	3.5	1161	1 NRDC RAT	Q9mu22 homo sapien
21	104.5	3.5	5596	1 MDN1 HUMAN	Q08198 halobacteri
22	104	3.5	852	1 CSG_FALHA	Q18346 equine hebp
23	103.5	3.5	273	1 IE68 HSVB4	Q40349 utillago ma
24	103	3.4	950	1 URBI_USTMA	Q07897 rattus nov
25	102.5	3.4	2124	1 PGCA RAT	Q9Y6A5 homo sapien
26	102	3.4	838	1 TAC3 HUMAN	Q39520 saccharomyc
27	102	3.4	1085	1 IFH1_YEAST	Q48434 gallus gall
28	101.5	3.4	494	1 SOX9 CHICK	Q81YJ3 homo sapien
29	101.5	3.4	562	1 STL1 HUMAN	Q95819 homo sapien
30	101.5	3.4	1239	1 M4K4 HUMAN	Q04538 t genome po
31	101	3.4	3415	1 POLG POWVL	Q60840 homo sapien
32	100.5	3.4	1966	1 CCAP_HUMAN	P15205 rattus nov
33	100.5	3.4	2459	1 MAPE_RAT	

ALIGNMENTS

RESULT 1
NCR2_HUMAN
ID NCR2_HUMAN STANDARD; PRT; 2517 AA.
AC Q9Y618; Q00613; Q15416; Q13354; Q9Y5U0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nuclear receptor co-repressor 2 (N-COR2) (Silencing mediator of
DE retinoic acid and thyroid hormone receptor) (SMRT) (SMRTE) (Thyroid-,
DE retinoic-acid-receptor-associated co-repressor) (T3 receptor-
DE associating factor) (TRAC) (CTG repeat protein 26).
GN NCR2 OR CTG26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM SMRT).
RX TISSUE=Pituitary;
RC MEDLINE=99178941; PubMed=10077563;
RA Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
RT "Unique forms of human and mouse nuclear receptor corepressor SMRT";
RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SMRT).
RX TISSUE=Cervical adenocarcinoma;
RC MEDLINE=9919215; PubMed=10097068;
RA Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;
RT "SMRte, a silencing mediator for retinoid and thyroid hormone
RT receptors-extended isoform that is more related to the nuclear
RT receptor corepressor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
RN [3]
RP SEQUENCE OF 1023-2517 FROM N.A.
RX TISSUE=Cervical adenocarcinoma;
RC MEDLINE=96008552; PubMed=7566127;
RA Chen J.D., Evans R.M.;
RT "A transcriptional co-repressor that interacts with nuclear hormone
RT receptors.";
RL Nature 377:454-457(1995).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM TPAC-1).
RX TISSUE=Fetal liver;
RC MEDLINE=96408715; PubMed=8813722;
RA Sande S., Privalsky M.L.;
RT "Identification of TRACs (T3 receptor-associating cofactors), a family
RT of cofactors that associate with, and modulate the activity of,
RT nuclear hormone receptors.";
RL Mol. Endocrinol. 10:813-825(1996).
RN [5]
RP SEQUENCE OF 428-613 FROM N.A.
RX TISSUE=Brain cortex;
RC MEDLINE=97369492; PubMed=9225980;
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,
RA Breschel T.S., Steine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT "cDNAs with long CAG trinucleotide repeats from human brain.";

34 99.5 3.3 426 1 ICSB_HUMAN Q02556 homo sapien
35 99.5 3.3 1601 1 PG2_HUMAN Q8tau7 homo sapien
36 99.5 3.3 2364 1 PGCA_BOVIN Q1608 bos taurus
37 99 3.3 1058 1 BUB1_MOUSE Q08901 mus musculus
38 98.5 3.3 456 1 EYZ2_CHLRE Q08355 chlamydomon
39 98.5 3.3 695 1 ARP_EUGGR Q04732 euglena gra
40 98 3.3 461 1 SX10_CHICK Q9W757 gallus gra
41 98 3.3 2468 1 MAPB_HUMAN Q46821 homo sapien
42 98 3.3 4625 1 DV1A_CHLRE Q9smh3 chlamydomon
43 97.5 3.3 536 1 CH62_CHLCV P58698 chlamydomon
44 97.5 3.3 1131 1 MYPF_CHICK P16419 gallus gall
45 97 3.2 528 1 RPSA_MYCTU Q60162 mycobacteri

RL Hum. Genet. 100:114-122(1997).
 RP [6]
 RP INTERACTION WITH MINT.
 RX MEDLINE=21231190; PubMed=11331609;
 RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
 RA Hon M., Evans R.M.,
 RA "Sharp, an inducible cofactor that integrates nuclear receptor
 RT repression and activation.";
 RL Genes Dev. 15:1140-1151(2001).
 RN [7]
 RP INTERACTION WITH HDAC10.
 RP MEDLINE=21839031; PubMed=11739383;
 RA Fischer D.D., Cai R., Bhatia U., Asselbergs F.A.M., Song C., Terry R.,
 RA Trognani N., Widmer R., Acadja P., Cohen D.,
 RT "Isolation and characterization of a novel class II histone
 RT deacetylase, HDAC10.";
 RL J. Biol. Chem. 277:6656-6666(2002).
 CC -1- FUNCTION: Mediates the transcriptional repression activity of some
 CC nuclear receptors by promoting chromatin condensation, thus
 CC preventing access of the basal transcription.
 CC -1- SUBUNIT: Interacts with HDAC7 (By similarity). Forms a large
 CC corepressor complex that contains SIN3A/B and histone deacetylases
 CC HDAC1 and HDAC2. This complex associates with the thyroid (TR) and
 CC the retinoid acid receptors (RAR) in the absence of ligand, and
 CC may stabilize their interaction with TFIIB. The SRMT isoform
 CC interacts with HDAC10. Interacts with MINT.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=SMRT; Synonyms=TRAC-2;
 CC IsoId=Q9Y618-1; Sequences=Displayed;
 CC Name=TRAC-1;
 CC IsoId=Q9Y618-2; Sequences=VSP_003412, VSP_003413;
 CC Note=Contains only the C-terminal receptor-interacting domain
 CC and acts as an antirepressor;
 CC -1- TISSUE SPECIFICITY: Ubiquitous. High levels of expression are
 CC detected in lung, spleen and brain.
 CC -1- INDUCTION: Regulated during cell cycle progression.
 CC -1- DOMAIN: The N-terminal region contains repression functions that
 CC are divided into three independent repression domains (RD1, RD2
 CC and RD3). The C-terminal region contains the nuclear receptor-
 CC interacting domains that are divided in two separate interaction
 CC domains (ID1 and ID2).
 CC -1- DOMAIN: The two interaction domains (ID) contain a conserved
 CC sequence referred to as the CORNR box. This motif is required and
 CC sufficient to permit binding to unliganded TR and RARs. Sequences
 CC flanking the CORNR box determine nuclear hormone receptor
 CC specificity.
 CC -1- SIMILARITY: Contains 1 SANT-A domain.
 CC -1- SIMILARITY: Contains 1 Myb-like domain.
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF113003; AAC20946.1; -;
 DR EMBL; AF125672; AAC22973.1; -;
 DR EMBL; U37146; AAC50236.1; -;
 DR EMBL; S83390; AAB50847.1; -;
 DR EMBL; U80750; AAB91446.1; -;
 DR F01; S60255; S60255.
 DR PDB; 1KKQ; 27-FEB-02.
 DR TRANSFAC; T04689;
 DR Genew; HGNC:7673; NCOR2.
 DR MIM; 600848; -;
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0003714; F:transcription co-repressor activity; TAS.

DR InterPro; IPR001005; Myb DNA binding.
 DR Pfam; PF00249; myb DNA-binding; 2.
 DR SMART; SM00717; SANT; 2.
 DR PROSITE; PS0090; MYB_3; 1.
 KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
 KW Coiled coil; Alternative splicing; 3D-structure.
 FT DOMAIN 174 215
 FT COILED COIL (POTENTIAL).
 FT INTERACTION WITH SIN3A/B (BY SIMILARITY).
 FT SANT-A (POTENTIAL).
 FT MYB.
 FT COILED COIL (POTENTIAL).
 FT PRO-RICH.
 FT CORNR BOX OF ID1.
 FT CORNR BOX OF ID2.
 FT POLY-GLN.
 FT POLY-LYS.
 FT POLY-PRO.
 FT POLY-PRO.
 FT POLY-GLY.
 FT POLY-PRO.
 FT Missing (in isoform TRAC-1).
 FT /FtId=VSP_003412.
 FT Missing (in isoform TRAC-1).
 FT /FtId=VSP_003413.
 FT L -> P (IN REF. 2).
 FT K -> E (IN REF. 2).
 FT L -> W (IN REF. 2).
 FT MISSING (IN REF. 2).
 FT A -> P (IN REF. 2).
 FT SS -> EF (IN REF. 5).
 FT S -> T (IN REF. 2).
 FT MISSING (IN REF. 2).
 FT KRISAPIEP -> PEDIPAPTES (IN REF. 2).
 FT G -> L (IN REF. 2).
 FT S -> F (IN REF. 2).
 FT A -> S (IN REF. 2).
 FT G -> R (IN REF. 2).
 FT SRSPAPPA -> MEAWDHP (IN REF. 3).
 FT A -> AEKPVFFPA (IN REF. 2).
 FT K -> T (IN REF. 4).
 FT P -> A (IN REF. 4).
 FT SEQUENCE 2517 AA; F5905C01761258C0 CRC64;
 Query Match 4.1%; Score 121.5; DB 1; Length 2517;
 Best Local Similarity 21.3%; Pred. No. 2.4;
 Matches 81; Conservative 48; Mismatches 132; Indels 119; Gaps 17;
 Qy 105 RWRLLCAEV--KECWVHARTHLHSGSLWELLYQHSVRLKHKRRRPRPFVGENSDS 162
 Db 632 RNWSAIARMVGSKTQSCNFKYENYKRNQNLDEILOQHLKMKERNARKKKAPAAAS 691
 Qy 163 SEEDHPAFCDVPVYQ-----TGAEEF-----DSGDE-----GPSTRHSAS 197
 Db 692 EEAAFP-----PVVEEMEASGVSGNEEMVEEAEALHASGNEVPRGECGPATVNNSS 746
 Qy 198 GVQ-----PVDDANADFGSGDEGFSTRHSDSOPP-----PADETT--- 233
 Db 747 DTESISPHTEAAKDGQNGKPKPATLGADGPPGPTPPRTSRAPISPTPASEATGAP 806
 Qy 234 -----VHTDNVEDDLTLDKESACALMYHVQEMDMMLAMCDEDFLL 278
 Db 807 TPTPAPPSPSAPPVVPKKEE-----ETAAAPVVEEGE-----QKPPAAEELAVDT 855
 Qy 279 GIPEDVIAT-----QPGGDTDASGVVTEGSIASAAGVAGVEDVYLAGALEAQNVA 329
 Db 856 GKAEFPVKSECTEAEEGPAKGDAAEATAEGALKAEKKEGG-----SGEATTTAKS 909
 Qy 330 G-----EYVLEISDEHVDGAG-----LPPASRRRPFVVEFLWDGDPHRRHP--TTRII 377
 Db 910 GAPQDSSTATCSADEVDEAEAGDKNRLSPRPSLITPTGD-----PRANASPKQLDL 963
 Qy 378 RHRKLSAYRVARPPVMLT 397

Db 964 KOLKORAA-----AIPPIQVT 979

RESULT 2

FTPX_RAT STANDARD; PRT; 1004 AA.

AC Q63475;

DT 01-NOV-1997 (Rel. 35, Last created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Receptor-type protein-tyrosine phosphatase N2 precursor (EC 3.1.3.48)

DE (R-PTP-N2) (PTP NE-6) (PTPNE6) (Phogrin).

GN PTPN2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=New England Deaconess Hospital; TISSUE=Insulinoma;

RA Wasmeyer C., Hutton J.C.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley;

RA Fitzgerald L.R., Walton K.M., Dixon J.E., Largent B.L.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Implicated in development of nervous system and pancreatic endocrine cells.

CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein tyrosine + phosphate.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).

CC -1- TISSUE SPECIFICITY: Pancreas.

CC -1- PTM: Appears to undergo multiple proteolytic cleavage at consecutive basic residues (By similarity).

CC -1- SIMILARITY: Contains 1 protein-tyrosine phosphatase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z50735; CAA90600.1; -

DR EMBL; U73458; AAC08036.1; -

DR HSP; P18052; IYFO.

DR InterPro; IPR000387; TYR phosphatase.

DR InterPro; IPR000242; Tyr_PP.

DR Pfam; PF00102; Y_phosphatase; 1.

DR PRINTS; PR00700; PRTYPHPTASE.

DR SMART; SM00194; PTPc; 1.

DR PROSITE; PS00383; TYR_PHOSPHATASE 1; 1.

DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.

DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.

KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane.

FT SIGNAL 1

FT CHAIN 27

FT RECEPTOR-TYPE PROTEIN-TYROSINE PHOSPHATASE N2.

FT PHOSPHATASE N2.

FT EXTRACELLULAR (POTENTIAL).

FT POTENTIAL.

FT CYTOPLASMIC (POTENTIAL).

FT DOMAIN 28 603

FT TRANSMEM 604 624

FT DOMAIN 625 1004

FT DOMAIN 759 993

FT ACT_SITE 934 934

FT SITE 415 416

FT CLEAVAGE SITE (BY SIMILARITY).

FT CARBOHYD 553 553

FT N-LINKED (GLCNAC...) (POTENTIAL).

FT SEQUENCE 1004 AA; 111863 MW; A73929E11B486F2 CRC64;

Query Match 3.9%; Score 118; DB 1; Length 1004;

Best Local Similarity 21.8%; Pred.No. 1.3; Indels 198; Gaps 27;

Matches 115; Conservative 60; Mismatches 159;

QY 153 RPFVGENSDSEEDHPAFCDYVPTOTGAESDSDGDEGSTR-----HSASGVQPYD- 203

Db 177 RPKVGE--DSSPENILTY-----VAHTSALTYP-----PATRVKYPDNLLRPLSLRQPDEL 225

QY 204 ----DANADS-----PGSGDEGSPSTRHSDSQ----- 226

Db 226 SPKVDSDIDKQKLI AALGAYTAQRPPGEND--PEPRYL VHSMPRAPFAAPALSQRNPL 283

QY 227 PPADETTHVDNVEDDLILLKESACALMYHVGQ--ENDMLMRAMCDEDLFDLLG-IPED 283

Db 284 PGDS-----KDSLSMGDDTLRLSLKLOLQQAQAEVDRLGSLKLEBQADSIAQAIQSD 335

QY 284 VIATSPQGGDTDASGVWTEGSIATASAVGAGVEDVYLALAEANQVAGVYVLEISDE---E 340

Db 336 PVEGSO-----ZSHGRGAS-----CQLREQADAPEMLQ--DRLIPE 370

QY 341 VDDGAGLPPASRRRPPVGVFEFWDDGP-----RRHERP-----TTR 375

Db 371 VDDPAAYKEVSRSLSPKGLDGLKDHGSPILLPEAPLLEKSSRAEMKKSEQPEEVSSEETA 430

QY 376 RIHRKLR-----AYRVA-----RPPVWITDRLGVVFYGRPA--NSL 414

Db 431 GVEHVKSRTYSKDLLERKPNSEPOFWRLEDFONRAPEVWEDEQNKLKLAAGPPSGGQL 490

QY 415 EV-----ERKVFILCSQNPLA-----DISHSLHSRKLRLVLPKP-----DDNNT 455

Db 491 EVQPEEEOQGVILGNPNPLSPKGLQMLDEVAH-----LLRVPSSFFADVKVL 539

QY 456 GPGDVNLLAALVRSF-----ASGLVIVSLRSGIYVKNLCKSTVLVHG 497

Db 540 GPAVIFKVSANIQNNTTADVTKAAVDNKNDELEKATGLTI--LQSGIRPKGLKLLPHPEE 597

QY 498 NNPKPKGVICGLSSRAVLDFENVAQ-----YRIQHEHIKKTTFVIGGDPTS 545

Db 598 QEDSKFVILFSLACILAVLLASSLAYCLRHNSHYKLEKLSGLGADPSA 649

RESULT 3

SDX10 MOUSE STANDARD; PRT; 466 AA.

ID SDX10 MOUSE

AC Q04888; O08518; O09141; O54856; P70416;

DT 01-JUN-1994 (Rel. 29, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Transcription factor SOX-10 (SOX-21) (Transcription factor SOX-M).

GN SOX10 OR SOX-10 OR SOX21.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RC MEDLINE=98085871; PubMed=9425902;

RA Southard-Smith E.M., Kos L., Pavan W.J.;

FT "Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model.";

RL Nat. Genet. 18:60-64 (1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RC MEDLINE=98226785; PubMed=9560246;

RA Herbarth B., Pingault V., Bondurant N., Kuhlbrodt K.,

RA Hermans-Borgmeyer I., Puliti A., Lemort N., Goossens M., Wegner M.;

RT "Mutation of the hry-related Sox10 gene in dominant megacolon, a mouse model for human Hirschsprung disease.";

RL Proc. Natl. Acad. Sci. U.S.A. 95:5161-5165 (1998).

RN [3]

RP SEQUENCE FROM N.A.

RC MEDLINE=98430992; PubMed=9760192;

RA Pusch C., Huestert E., Pfeifer D., Sudbeck P., Kist R., Roe B.,

RA Wang Z., Balling R., Blin N., Scherer G.;

RT "The SOX10/Sox10 gene from human and mouse: sequence, expression, and

RT transactivation by the encoded HMG domain transcription factor.";
 RL Hum. Genet. 103:115-123(1998).

[4]

RP SEQUENCE FROM N.A.

RC STRAIN=C3H/HeN;

RX MEDLINE=97179197; PubMed=9027493;

RA Tani M., Shindo-Okada N., Hashimoto Y., Shiroishi T.,

RA Takenoshita S., Nagamachi Y., Yokota J.,

RT "Isolation of a novel Sry-related gene that is expressed in high-

RL metastatic K-1735 murine melanoma cells.";

RT Genomics 39:30-37(1997).

[5]

RN SEQUENCE FROM N.A.

RP MEDLINE=22388257; PubMed=12477932;

RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusich K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.B.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S., Krzywinski M.I., Skalska U., Smalls D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,

RT "Generation and initial analysis of more than 15,000 full-length

RL human and mouse cDNA sequences.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[6]

RN SEQUENCE OF 112-167 FROM N.A.

RP MEDLINE=93181275; PubMed=8441686;

RA Wright E.M., Snopce B., Koopman P.,

RT "Seven new members of the Sox gene family expressed during mouse

RT development.";

RT Nucleic Acids Res. 21:744-744(1993).

[7]

RN SEQUENCE OF 114-167 FROM N.A.

RP STRAIN=Swiss Webster; TISSUE=Embryonic tooth;

RX MEDLINE=97079683; PubMed=8921394;

RA Stock D.W., Buchanan A.V., Zhao Z., Weiss K.M.,

RT "Numerous members of the Sox family of HMG box-containing genes are

RT expressed in developing mouse teeth.";

RT Genomics 37:234-237(1996).

CC -!- FUNCTION: Transcription factor that seems to function

CC synergistically with the POU domain protein TST-1/OCT6/SCIP. Could

CC confer cell specificity to the function of other transcription

CC factors in developing and mature glia (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- DISEASE: Defects in SOX10 are the cause of the mouse mutant

CC dominant megacolon (dom). While dom/+ heterozygous mice display

CC regional deficiencies of neural crest-derived enteric ganglia in

CC the distal colon, dom/dom homozygous animals are embryonic lethal.

CC -!- SIMILARITY: Contains 1 HMG box domain.

CC -!- CAUTION: Ref 4 sequence differs from that shown due to

CC frameshifts in positions 16 and 392.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF017182; AAB99738.1; ALT_INIT.

DR EMBL; AF047043; AAC24564.1;

DR EMBL; U66141; AAB49282.1; ALT_FRAME.

DR EMBL; BC018551; AAH18551.1; --
 DR EMBL; Z18959; CAA79484.1; --
 DR EMBL; U70441; AAC52859.1; --
 DR PIR; S30242; S30242.
 DR HSP; P48436; ISX9.
 DR MGD; MG1:98358; Sox10.
 DR GO; GO:0030154; P:cell differentiation; IMP.
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 1.
 DR SMART; SM00398; HMG; 1.
 DR PROSITE; PS01118; HMG_BOX_2; 1.
 KM transcription regulation; DNA-binding; Nuclear protein.
 FT DOMAIN 35 41 POLY-GLY.
 FT DNA BIND 104 172 HMG BOX.
 FT VARIANT 11 11 E -> V.
 FT CONFLICT 41 41 G -> V (IN REF. 4).
 FT CONFLICT 119 119 R -> P (IN REF. 6).
 FT CONFLICT 357 357 K -> E (IN REF. 4).
 SQ SEQUENCE 466 AA; 49949 MW; 84862547F039E71F CRC64;

Query Match 3.9%; Score 116.5; DB 1; Length 466;

Best Local Similarity 19.7%; Pred. No. 0.58;

Matches 81; Conservative 46; Mismatches 147; Indels 137; Gaps 19;

QY 119 WCYHAR-----THLHSG-----SSLWEILYQHSVRLEKRRPRRRFVGENSD---S 162

DB 114 WQAARRLADQYPHLHNAELSKTLKRLINLNSD-----KRPFIIEAERLRMQ 163

QY 163 SEDHPAFCDVPVTO-----TGAESEDSGDEGPST-----RHSASGVQP 201

DB 164 HKKDHPDYQYRRRRKNGKAAQGEAECPGAEAGGAAAIQAHYKSAHLDRHPPEG-SP 222

QY 202 VDDANADSPGSGDEGSPSTRHSQPPPADETV-----HTDNVED 241

DB 223 MSDGNPEHSPGSHGPPT-----PPTTPEKLOSGKADPKDGRSLGEGKPHIDFGNV 276

QY 242 DLTLLDKESACALMYHVQGMMLMRAMCDEDLFLLGIPEDVIATISQGGTDSAGVVT 301

DB 277 DGEISHE-----VMSNMETFDVTELDQYLPNGHFG-----H 309

QY 302 EGSIAASAVGAGVEDVYLAGALEAQNAGEYVLEISDEVEDDAGLPPAS-----RRRPV 356

DB 310 VGSYSAAGYGLG-----SALAVASGHSAWISK-----PPGVALPTVSPPGVDAKAQV 356

QY 357 VGEFLWDDGPRRH-ERPTTRRTRHRKLR-----SAYYVARPPVMTDLRGVEVFEGRP 410

DB 357 KTTCTPGQPPHYTDQPTSQIATYLSLPHYGSAPPSISRFQFDYSDHQPSGP-YIGHA 415

QY 411 AMSLEY-----ERKVFILCSQ-NPLADISHSLSRKGLRLLPKP 450

DB 416 GQASGLYSAFSYMGPSORPLYTAISDPSPSGQSHSPHWEQPVYTTLSRP 466

RESULT 4

Y456_CHLTR

ID Y456_CHLTR STANDARD; PRT; 1005 AA.

AC 084462;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Protein CT456 precursor.

GN CT456.

OS Chlamydia trachomatis.

OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=813;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=D/UV-3/Cx;

RX MEDLINE=99000809; PubMed=9784136;

RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,

RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,

RA Davis R.W.;

RT "Genome sequence of an obligate intracellular pathogen of humans:


```

FT CONFLICT 554 554 V -> M (IN REF. 1).
FT CONFLICT 575 581 FRYLEP -> LPLCARA (IN REF. 1).
SQ SEQUENCE 1007 AA; 105398 MW; 24B7C4ADE852B96C CRC64;

Query Match
Best Local Similarity 19.4%; Score 113; DB 1; Length 1007;
Matches 111; Conservative 61; Mismatches 171; Indels 228; Gaps 25;

QY 146 KRRRRPRR-----PFVGENSDSEEDHPAFCDVFTVQ-----TGAESD----- 184
DQ 4 RKQRQQLISDCGEPFASENGDAEEDHPQVCAKCCQAQFTDTEFLAHQACSTDPVPM 63
QY 185 ----SGDEGSTRHSAGVOP-----VDDANADSPGSGDE-----GPSTRHSDQ- 225
DQ 64 VIIGQENFNSSASSEPPGEGHNPQVMDTHSPGSSVPTDPTWGERGEESSG 123
QY 226 -----PPADETTVHTDNVE----- 240
DQ 124 HFLVAATGTAAGGGGLILASPKLGATPLPPESTPAPPPPPPPPPGSGSHLNPLIL 183
QY 241 DDLTLLDK-----ESAC---ALMYHVGO-----EMDMLMEAMCDEDLFDLLG 279
DQ 184 EELRVLQQRQIHMOMTEOICQVLLGSLGTGAPASPSELPTGTASSTKPLPLFS 243
QY 280 IPEDV-----IATSPQSGDT-----DASGV----- 299
DQ 244 PIKPVQTSKTLASSSSSSSSSSGAETPKQAFPHLYPLGSHQHPFSGAGVGRSHKPTAPS 303
QY 300 -----VTGSIASAVGA--GVEDVYLAGALEAONVAGEYVLEISD 338
DQ 304 PALPGSTDQLIASPHLAFPTTGLLAAQCLGAARLEATASPLGLKPKNGSG-----ELSY 359
QY 339 EEVDGAGLPPASRRPVPVGEFLWDGP-----RHH--ERP-----TTR---RIR 378
DQ 360 GEWVGLEKPGGGRHKRCFCFAKVGSDSALQIHLRSHTGERPYKCVGNRTTRGNLKVH 419
QY 379 HKLRSAIYRVARPPVMTDRIGVEFYFGRP-ANSL-----VERKVFIL 423
DQ 420 FRRHREKYPHVQNPVPPVPEHLDYVITSSGLPYGMSVPEKAEBAEATPGGQVERKPLVA 479
QY 424 CSQNPADISHSCLSRKSG-----LRVLLPK-PDNNNGTGP-DVNLLAAV 466
DQ 480 SITALSATESLILSTISATATAPGLPARKVKVMAKVPKADENPTFPGSEGAISGV 539
QY 467 LRSFAS-----GLVIVLSRSGIYVKNLCKST 492
DQ 540 ABSSTATRMLQSLKVLTSPLSWALLTNHFKST 570

RESULT 6
SK10_HUMAN STANDARD; PRT; 466 AA.
AC P5693;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor SOX-10.
GN SOX10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT WS4 LEU-ARG-161 INS.
RX MEDLINE=98122581; PubMed=9462749;
RA Pingault V., Bondurand N., Kuhlbrodt K., Goerich D.E., Prehu M.O.,
RA Puliti A., Herbarth B., Hermans-Borgmeyer I., Legius E., Matthijs G.,
RA Amiel J., Lyonnet S., Ceccherini I., Romeo G., Clayton Smith J.,
RA Read A.P., Wegner M., Goossens M.;
RT "SOX10 mutations in patients with Waardenburg-Hirschsprung disease.";
RL Nat. Genet. 18:171-173(1998).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=98430992; PubMed=9760192;
RA Pusch C., Rustert E., Pfeifer D., Sudbeck P., Kist R., Roe B.,
RA Wang Z., Bailing R., Blin N., Scherer G.;
RT "The SOX10/Sox10 gene from human and mouse: sequence, expression, and
transactivation by the encoded HMG domain transcription factor.";
RL Hum. Genet. 103:115-123(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgen A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Dami P.D., Dockree K., Dodsworth S.J., Cox A.V., Davis J., Dawson E.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.W., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay Y., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuayama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Lareille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker J.A., Wamsley A., Weidmann P., Pepin K., Nelson J.,
RA Kori I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saatta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edlmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Serousi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tiliahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [4]
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta, and Skin;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

```


"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[5]
CHARACTERIZATION.
RX MEDLINE=98389729; PubMed=9722528;
RA Kuhlbrodt K., Schmidt C., Sock E., Pingault V., Bondurand N.,
RA Goossens M., Wegner M.;
RA "Functional analysis of Sox10 mutations found in human Waardenburg-
RT Hirschsprung patients.";
RT J. Biol. Chem. 273:23033-23038(1998).
RN [6]
RN NUCLEOCYTOPLASMIC SHUTTLING.
RX MEDLINE=22133596; PubMed=12138193;
RA Rehberg S., Lischka P., Glaser G., Stamminger T., Wegner M.,
RA Rosorius O.;
RA "Sox10 is an active nucleocytoplasmic shuttle protein, and shuttling
RT is crucial for Sox10-mediated transactivation.";
RT Mol. Cell. Biol. 22:5826-5834(2002).
CC -I- FUNCTION: Transcription factor that seems to function
CC synergistically with the POU domain protein TST-1/OC76/SCIP. Could
CC confer cell specificity to the function of other transcription
CC factors in developing and mature glia (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC -I- TISSUE SPECIFICITY: Expressed in fetal brain and in adult brain,
CC heart, small intestine and colon.
CC -I- DISEASE: Defects in SOX10 are a cause of Waardenburg syndrome
CC type IV (WS4) [MIM:277580]; also known as Waardenburg-Shan
CC syndrome. WS4 is characterized by the association of Waardenburg
CC features (depigmentation and deafness) and the absence of enteric
CC ganglia in the distal part of the intestine (Hirschsprung
CC disease).
CC -I- SIMILARITY: Contains 1 HMG box domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ001183; CRA04576.1; -
DR EMBL; AL031587; CAB62982.1; -
DR EMBL; BC002824; AAH02824.1; -
DR EMBL; BC007345; AAH07595.1; -
DR HSSP; P48436; ILSX9.
DR TRANSFAC; T04918; -
DR Genew; HGNC:11190; SOX10.
DR MIM; 602229; -
DR MIM; 277580; -
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR GO; GO:0003713; F:transcription co-activator activity; TAS.
DR GO; GO:0007345; F:embryogenesis and morphogenesis; TAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SMC0398; HMG; 1.
DR PROSITE; PS0118; HMG_BOX_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW Disease mutation; Hirschsprung disease; Deafness;
FT DOMAIN 134 145 NUCLEAR EXPORT SIGNAL.
FT FT 135 41 POLY-GLY.
FT DOMAIN 104 172 HMG BOX.
FT DNA_BIND 161 161
FT VARIANT 161 161
FT FT R -> RLR (in WS4).
FT FT /FTID=VAR_003743.
SQ SEQUENCE 466 AA; 49911 MW; FAAIECI08D4DE6A1 CRC64;

Query Match 3.88; Score 112.5; DB 1; Length 466;
Best Local Similarity 19.58; Pred.No.1.1;
Matches 80; Conservative 46; Mismatches 148; Indels 137; Gaps 19;

QY 119 WCYHAR-----THLHSG-----SSLWEILYQHVSVRLEKRRRRRRPVPVGENSD----S 162

FT DNA BIND 104 172 HMG BOX.
 SQ SEQUENCE 466 AA; 50039 MW; 094B0C5E42D83454 CRC64;
 Query Match 3.8%; Score 112.5; DB 1; Length 466;
 Best Local Similarity 20.3%; Pred. No. 1.1;
 Matches 83; Conservative 50; Mismatches 143; Indels 133; Gaps 21;
 QY 119 WCWHAR-----THLHSG-----SSLWEILYQHSVLEKRRRPRPPFVGENSD--S 162
 DB 114 WAQAARRKLADQYPLHNAELSKTLGLWRLNESD-----KEPFTIEAERLRMQ 163
 QY 163 SEEDHPACDVPTQTG-----AESDSG-----DEGPST-----RHASGVQP 201
 DB 164 HKXDHDPYKYPRRRKNGKAQGAECFGETDQGAAGAAIOAHYKSAHLDRHPPEEG-SP 222
 QY 202 VDNADSPGSGDEGFSRHSQPPADETTV-----HTDNVED 241
 DB 223 MSDGNPEHPSGSHQPP-----PPTPKTELQSGKADPKRDSIGSGCKPHIDFGV 276
 QY 242 DUTLKDSEACALMYHVGEMDMMLRACDEDLFDLIGIPEDVIATSQPG--GPTDASGV 299
 DB 277 DIGEISHE-----VMSNMETFDVTDLQYLPNGHPGHVGSYSAAGY 318
 QY 300 VTEGSIASAVGAGVEDVYLAGALEAQNAGVEYVLEISDEVDGAGLPPA-SRRRPVVG 358
 DB 319 ---GLSSALAVSG-----HSWISKPPGVALPTV-----SPPAVDRAKQVKT 358
 QY 359 EFLWDGPRRH-ERPTTRIRRRKLRL-----SAYRVARPPVMTDRLGVEVFYGRFAM 412
 DB 359 ETTGQGPPEHYTDQSTQSLAYTSLSLPHYSARFPISRQPDYSDHQSPG-YYGHAGQ 417
 QY 413 SLEV-----ERKVFILCSQ-NPLADISHSLHRSKGLRVLKXP 450
 DB 418 ASGLYSAFSYMGPSORPLTYAISDPSPGQSHSPTHWEQVYVYTLGRP 466

RESULT 8

ID NRDC MOUSE STANDARD; PRT; 1161 AA.
 AC Q8BNG1;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Nardilysin precursor (EC 3.4.24.61) (N-arginine dibasic convertase)
 DE (NRD convertase) (NRD-C).
 GN NRDI...musculus (Mouse).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 EX MEDLINE=23254683; PubMed=12468851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gajobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Giesi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Saitano R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,

Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu M.,
 Hironaka-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukawa S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sakai D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Kidney;
 EX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zesberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnarch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -|- FUNCTION: Cleaves peptide substrates on the N-terminus of arginine
 CC residues in dibasic pairs.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of polypeptides, preferably at
 CC Xaa-|-arg-Lys, and less commonly at Arg-|-arg-Xaa, in which Xaa is
 CC not Arg or Lys.
 CC -|- COFACTOR: Zinc (Probable).
 CC -|- SIMILARITY: Belongs to peptidase family M16.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AK031548; BAC27445.1; -;
 CC EMBL; BC036128; AH36128.1; -;
 CC MGD; MG1:1201386; Nrdl.
 DR InterPro; IPR001431; Peptidase_M16.
 DR InterPro; IPR007863; Peptidase_M16_C.
 DR Pfam; PF00675; Peptidase_M16; 1.
 DR Pfam; PF05193; Peptidase_M16; 2.
 DR PROSITE; PS00143; INSULINASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1161
 FT POTENTIAL
 FT NARDILYSIN.
 FT DOMAIN 139 209 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 140 151 POLY-GLU.
 FT DOMAIN 152 165 POLY-ASP.
 FT DOMAIN 185 190 POLY-ASP.
 FT DOMAIN 193 196 POLY-ASP.
 FT METAL 244 244 ZINC (BY SIMILARITY).
 FT ACT_SITE 247 247 BY SIMILARITY.
 FT METAL 248 248 ZINC (BY SIMILARITY).
 FT METAL 325 325 ZINC (BY SIMILARITY).
 SQ SEQUENCE 1161 AA; 132890 MW; 21334221632A5122 CRC64;
 Query Match 3.7%; Score 111; DB 1; Length 1161;

Best Local Similarity 19.0%; Pred. No. 4.8;
Matches 118; Conservative 64; Mismatches 190; Indels 250; Gaps 29;
QY 77 AFGKICITTSRLRLPGEVEVVGINCRWRLLCAEVECW--WCVHARTHLSGSSLW 134
Db 6 AVAAVCVTGRKLRCEAGRELTALGRIBARG--LCEESSKPPFTLWPGRNKAKSTCSCP 62
QY 135 BILYQHSVRLEKRRRRPRRPVGEN-----SDSSEDEHPAFCDVPVTO----- 177
Db 63 DL-----QNGQDLGSGRLARLGADESEEGSFSNVGDPILIKSPDPKQ 109
QY 178 -----TG-----ASESDGSGPSTHUSAG 198
Db 110 YRIKLQNGLOALLISLDSNVGKTNATDEEEEBEEDDDDDDDDDDDDDSGAE 169
QY 199 VQPVDDANADSPGSDGSPSTHSDSQPPADETVHTDNDVDDLTLL-----D 247
Db 170 IQDDDESGFDEEEDDDDDHDDDD-----LENEELEELEERVEARKTTE 219
QY 248 KESACALMYHVGQMDMLRAMDCEDLFDLLGI-----PE-----DVIATS 288
Db 220 QKSAALCVGVG-----SFADPD--DLPLGLAHFLEHVMVFGSLKYPDENGDAFLKK 269
QY 289 QPGGD--TASGVVTEGSAASAVGAGVEDVLAGAL-----EAQ 326
Db 270 HGGSDNASTDCERTVFO-----FDVQKYPKEALDRWAFFIHLMIRDAIDREVE 320
QY 327 NVAGEYVL-----EISDEVEDDAGLPPASRRRP--VVGEEFLWDDGPRRHERPTTRIR-RH 379
Db 321 AVDSEYQIARSDANRKEMLFG-----SLAPGHPMGKFFWNAETLKHEPKKNIDTH 374
QY 380 KLASAYRVARPPVMTDRLGVFVFGPAPMSLEVERK-----VFILCSQNPLA 430
Db 375 ARLEEFNWR-----YSAHYMTLVVQSKETLDTLEKWTVEIFSQIP-- 415
QY 431 DISHSLHSRKLRLVLPKPDNN--TGPGDVNLLAALVR----- 468
Db 416 -----NNG-----LPKNFSLHTDPTFPAPKLYRVVPTRKHALTITWALPPQQ 462
QY 469 -----SPASGLVIVSLRGI--YKVLCKSTVLYHGNPPKFGVIGLSSRAVLDPV 519
Db 463 HYRVKPLHYISLWVHGEGKGSILYLRKKWALALFGNGE-----TGFEQNSYVSF 515
QY 520 NVA-QYRIQGHKIKT--TVF 538
Db 516 SISITLDEGEYHFAVHTVF 537

RESULT 9
AEGP RAT STANDARD; PRT: 1216 AA.
AC Q63191;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Apical endosomal glycoprotein precursor.
OS Rattus norvegicus (Rat)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestinal epithelium;
RX MEDLINE=95130529; PubMed=7829488;
RA Speelman B.A., Allen K., Grounds T.L., Neutra M.R.,
RA Kirchhausen T., Wilson J.M.;
RT "Molecular characterization of an apical early endosomal glycoprotein
RT from developing rat intestinal epithelial cells."
RL J. Biol. Chem. 270:1583-1588(1995).
CC -I- FUNCTION: Probably involved in the sorting and selective transport
CC of receptors and ligands across polarized epithelia.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -I- TISSUE SPECIFICITY: Apical endosomal tubules of developing rat

CC intestinal epithelial cells.
CC -I- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC -I- SIMILARITY: Contains 6 MAM domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L37380; AAA65200.1; --
DR PIR; A55620; DR
DR InterPro; IPR002172; LDL receptor_A.
DR InterPro; IPR000998; MAM domain.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00629; MAM; 6.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 3.
DR SMART; SM00137; MAM; 3.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00668; LDLRA_2; 2.
DR PROSITE; PS00740; MAM_1; FALSE NEG.
DR PROSITE; PS00660; MAM_2; 6.
KW Signal; Transmembrane; Glycoprotein; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 1216
FT DOMAIN 22 1155
FT TRANSMEM 1156 1176
FT DOMAIN 1177 1216
FT DOMAIN 27 54
FT DOMAIN 62 224
FT DOMAIN 229 269
FT DOMAIN 268 427
FT DOMAIN 454 491
FT DOMAIN 452 647
FT DOMAIN 654 813
FT DOMAIN 812 973
FT DOMAIN 972 1142
FT DISULFID 231 243
FT DISULFID 238 256
FT DISULFID 250 267
FT DISULFID 456 467
FT DISULFID 463 480
FT DISULFID 474 489
FT CARBOHYD 205 205
FT CARBOHYD 291 291
FT CARBOHYD 341 341
FT CARBOHYD 368 368
FT CARBOHYD 639 639
FT CARBOHYD 839 839
SQ SEQUENCE 1216 AA; 133776 MW; 64F3B28A7B61BA2E CRC64;
Query Match 3.7%; Score 111; DB 1; Length 1216;
Best Local Similarity 20.4%; Pred. No. 5.2;
Matches 99; Conservative 49; Mismatches 154; Indels 184; Gaps 25;
QY 30 RSTGFFLAARNVDVPLDSLOFFIDFKRECLSKGLHPRDLGSPITAFGKICTTSRLR 89
Db 708 RQCGALLTRPQVPVP-----KECLSPWYH---LHGQI---GTLCLAMRR-- 748
QY 90 RLPGEVEVW---OGINCRRRLCAEVEKWCWCVHARTHLSGSSSLWEIL-----YQ 139
Db 749 ---EGEEDTLWSRSGTHGNRW-----QAWTIL---HOLQSTKYQLFEGRLDGYH 796
QY 140 HSVRLKRRPRPRFPVGENSDSSEEDHPAFCDVPVTOGTGAESDSDGSGPSTR-HSASG 198
Db 797 GTMGDDLVAVRPGPCWAAKRCFSFDS--CGF-----SPGDWGLWTRQNNASG 842
QY 199 VQP-----VDDANADSPGS---GDGSP-----STRHSDSQPPADETVHTDNDVDD 243
Db 843 IGPWGFWDHTTGTGAQGHYVVDVTPNLLPKGHVASLTSEHPP----- 886

```
QY 244 TLDDKESACALMYHVGQEMDMLRAMCDELDLILGIPEDVIATSQPGDITDASGVVT-- 301
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
897 --LSRPACLSFWYHLSFNPGLRVEVE-----STRQELSLSGHGGAFAWRLGSVNVQ 938
QY 302 ---EGSIAASAVGACVEDVYLAGALEAQNVAAGEVLEISDEVDGACGLPPAS----- 351
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
939 AEQAWKVFEAWASGVESHVNA-----LDDISLQDGPQCAQSGCDFESGL 983
QY 352 -----RRRPVVGFEFLWD-----DGPFRHRPTTRIRHRKLRSAYYRVARPPVMTDLGV 402
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
984 CGWSHLPEGLGGYSDWSSGATPSRYPRPS-----VDHIVGT 1021
QY 403 EVFYP-----GRPMSLEVRKVFILCSQNPDLADISHSL-----HRRKG 442
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1022 EAGHFAPFETSVLGPGGAALGSE-----PLPATAVSCLHFWYMGFFAHPYKG 1071
QY 443 -LRVLL 447
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1072 ELRVLL 1077

RESULT 10
SAL2 MOUSE STANDARD; PRT; 1004 AA.
AC Q90X96;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Sal-like protein 2 (spalt-like protein 2) (Msal-2).
GN SAL2 OR SAL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEOLINE=20060974; PubMed=10602995;
RA Kohlase J., Altmann M., Archangelo L., Dickens C., Engel W.;
RT "Genomic cloning, chromosomal mapping, and expression analysis of
RT Msal-2.";
RL Mamm. Genome 11:64-68(2000).
CC -!- FUNCTION: Probable transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Expressed throughout embryonic development. In
CC adult predominantly in brain.
CC -!- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
FT ZN_FING 908 930 C2H2-TYPE 6.
FT ZN_FING 937 961 C2H2-TYPE 7.
FT DOMAIN 135 139 POLY-GLY.
FT DOMAIN 158 170 POLY-PRO.
FT DOMAIN 256 264 POLY-SER.
FT DOMAIN 762 772 POLY-GLU.
FT DOMAIN 834 840 POLY-PRO.
SQ SEQUENCE 1004 AA; 104975 MW; E46DD2C9C26A4C92 CRC64;

Query Match 3.7%; Score 109.5; DB 1; Length 1004;
Best Local Similarity 19.3%; Pred. No. 5.1;
Matches 111; Conservative 59; Mismatches 167; Indels 239; Gaps 25;

QY 146 KRRRRPR-----PFVGENSDSEEDHPAPCDVPVVTGTAESE----- 183
Db 4 RKQRKQQLISDCGFSASENGHHPQVCAKCAQFSDPTBFLAHQNSCCTDPPVM 63
QY 184 --DSGDEGSTRHSASGVOP-----VDDANADSPGSGDEGP----- 217
Db 64 VIIGQENPSNSASAPRPEGHRSQVMDTSHSNPPDSGSGPDDPTWGPERRREESSG 123
QY 218 -----STRHSDSQPPFADETTVHTDNVE----- 240
Db 124 QFLVAATGTAAGGGGLILASPKLGATPLPPESTAPPPTPPPPPGVSGHNLNPLIL 183
QY 241 DLTLLDK-----ESAC---ALMYHVGQEMDM----- 264
Db 184 EELRVLQQRQIHQMOWTEICRQVLLGSLGTVGAPASPSKLPGTGAASSTKPLLPFS 243
QY 265 -----LMRAMCEDLFDLIGIPEDVIATSQPG- 291
Db 244 PIKPAQTGKTLASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 303
QY 292 ----GDTD---ASGVV-----TEGSIASAVGA--GVEDVYLAGALEAQNVAAGE--YVLEIS 337
Db 304 ALPGSTDLQIASPHLAFPGTTGLLAACCLGAARGLEMAAASGLLRKPKNGSGELGYEVS 363
QY 338 DEEVDDGAGLPPASRRR-----PVVGEFLWDGPRRH-----ERP-----TTR- 375
Db 364 SLE-----KPGGRHKCFCAKVFGS---DSALQIHLRSHTGERPYKNCVNCNFTTG 413
QY 376 --RIRHRKLRSAYYRVARPPVMTDRLGVEVYFGRP-AMSL------VER 418
Db 414 NLKVHFRHREKYPHVQMPHEVPEHLDYVITSSGLPYGMSVPEPEKABEEAGTGGGV 473
QY 419 KVFILCSQNPDLADISHCLSRKG-----LRVLLPKPD-DNNTGPG-DVN 461
Db 474 KELVASTTALSATESLTLTSTGTSTAVAPGLPTFNKFLVLMKAVEPKSKADENTP 533
QY 462 LLAAVLRSPAS-----GLVIVSLRSGIYVKNLCKST 492
Db 534 AIAGVADSGSATRMQLSKLVTSLPSWALLTNHLKST 569

RESULT 11
IRF7_HUMAN STANDARD; PRT; 503 AA.
ID IRF7_HUMAN 000331; 000332; 000333; 075924;
AC Q92985; 000331; 000332; 000333; 075924;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Interferon regulatory factor 7 (IRF-7).
GN IRF7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Spleen;
RA Grossman A., Nicholl J., Antonio L., Luethy R., Suggs S.,
RA Sutherland G.R., Mak T.W.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
```


DT	10-OCT-2003 (Rel. 42, Last sequence update)
DT	10-OCT-2003 (Rel. 42, Last annotation update)
DE	Calmin.
GN	CLNN.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxId=10090;
[1]	
RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4), SUBCELLULAR LOCATION,
RP	AND TISSUE SPECIFICITY.
RC	STRAIN=ICG; TISSUE=Brain, and Testis;
EX	MEDLINE=21280911; PubMed=11386753;
RA	Ishizaki Z., Takaiishi M., Furuta I., Huh N.-H.;
RT	"Calmin, a protein with calponin homology and transmembrane domains
RT	expressed in maturing spermatogenic cells.";
RL	Genomics 74:172-179(2001).
[2]	
RP	SEQUENCE OF 98-1052 FROM N.A. (ISOFORM 1).
RC	STRAIN=C57BL/6J; TISSUE=Testis;
RX	MEDLINE=22354683; PubMed=12466851;
RA	Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA	Nikaido I., Osato N., Saito R., Suzuki H., Yamanaoka I., Kiyosawa H.,
RA	Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA	Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA	Schrim L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA	Blake J.A., Bradt D., Brusic V., Chotinia C., Corbani L.E., Cousins S.
RA	Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA	Gasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA	Grimmond S., Gustinchich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA	Kanai A., Kawaji H., Kawasaki Y., Kedzierski R.M., King B.L.,
RA	Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA	Maglott D.R., Maltais L., Marchionni L., McKenzie L., Mikhi H.,
RA	Negashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA	Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA	Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA	Sandelin A., Schneider C., Temple C.A., Setou M., Shimada K.,
RA	Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA	Vermado R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA	Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA	Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA	Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA	Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA	Hara A., Hashizume W., Imctani G., Ishii Y., Itoh M., Kagawa I.,
RA	Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA	Yaenunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA	Barney E., Hayashizaki Y.;
RT	"Analysis of the mouse transcriptome based on functional annotation o
RT	60,770 full-length cDNAs.";
RL	Nature 420:563-573 (2002).
[3]	
RP	TISSUE SPECIFICITY.
RP	MEDLINE=22557181; PubMed=12670712;
RA	Takaishi M., Ishisaki Z., Yoshida T., Takata Y., Huh N.-H.;
RT	"Expression of calmin, a novel developmentally regulated brain
RT	protein with calponin-homology domains.";
RL	Brain Res. Mol. Brain Res. 112:145-152(2003).
CC	-I- SUBCELLULAR LOCATION: Type IV membrane protein (Potential).
CC	Cytoplasmic. Isoforms 1 and isoform 4 show a reticular pattern in
CC	the cytoplasm
CC	
CC	-I- ALTERNATIVE PRODUCTS:
CC	Event-Alternative splicing; Named isoforms=4;
CC	Name=1; Synonyms=Beta;
CC	IsoId=Q8C5W0-1; Sequences=Displayed;
CC	Name=2; Synonyms=Delta;
CC	IsoId=Q8C5W0-2; Sequences=VSP_007766, VSP_007767;
CC	Note=Lacks the transmembrane domain;
CC	Name=3; Synonyms=Gamma;
CC	IsoId=Q8C5W0-3; Sequences=VSP_007768, VSP_007769;
CC	Note=Lacks the transmembrane domain;
CC	Name=4; Synonyms=Alpha;
CC	IsoId=Q8C5W0-4; Sequences=VSP_007770;
CC	-I- TISSUE SPECIFICITY: Expressed in testis. Expressed during testis

maturation process and in maturing spermatids. In brain, it is expressed in neurons of the hippocampus, cerebral cortex, and thalamus, Purkinje cells, and also in the choroid plexus and ependymal cells. Expressed predominantly in dendrites and cell bodies of the neurons, but not in axons. The level of expression increases during the period of maturation of the mouse brain after birth.

CC -!- SIMILARITY: Contains 1 actin-binding domain.
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

EMBL; AB047978; BAB59009.1; -
 EMBL; AB059643; BAB59120.1; -
 EMBL; AB059644; BAB59121.1; -
 EMBL; AB059645; BAB59122.1; -
 EMBL; AB059646; BAB59123.1; -
 EMBL; AB059647; BAB59124.1; -
 EMBL; AB059648; BAB59125.1; -
 EMBL; AK077023; BAC36573.1; -
 MGD; MGI:2136957; Clmn.
 GO; GO:0005737; Cytoplasm; IDA.
 InterPro; IPR001589; Actbind actnin.
 InterPro; IPR001715; Calponin-like.
 Pfam; PF00307; CH; 2.
 SMART; SM00033; CH; 2.
 PROSITE; PS00019; ACTININ_1; 1.
 PROSITE; PS00020; ACTININ_2; 1.
 PROSITE; PS00021; CH; 2.
 KW Actin-binding; Repeat; Transmembrane; Alternative splicing.
 FT TRANSMEM 1027 1047 ANCHOR FOR TYPE IV MEMBRANE PROTEIN (POTENTIAL).
 FT DOMAIN 1 288 ACTIN-BINDING.
 FT DOMAIN 1 139 CH 1.
 FT DOMAIN 187 288 CH 2.
 FT VARSPLIC 922 927 NGSIDS -> TWIPFL (in isoform 2).
 FT VARSPLIC 928 1052 /FTid-VSP 007766.
 FT VARSPLIC 942 946 Missing (in isoform 2).
 FT VARSPLIC 947 1052 /FTid-VSP 007767.
 FT VARSPLIC 947 1052 DFLSY -> SFHLY (in isoform 3).
 FT VARSPLIC 947 1052 /FTid-VSP 007768.
 FT VARSPLIC 947 1052 Missing (in isoform 3).
 FT VARSPLIC 947 1052 /FTid-VSP 007769.
 FT VARSPLIC 966 996 Missing (in isoform 4).
 FT CONFLICT 170 170 /FTid-VSP 007770. 2).
 FT CONFLICT 1012 1012 MISSING (IN REF. 2).
 FT CONFLICT 1012 1012 S -> R (IN REF. 2).
 SQ SEQUENCE 1052 AA; 117226 MW; ASAD7D3FF99A6EB6 CRC64;

Query Match 3.6%; Score 107; DB 1; Length 1052;
 Best Local Similarity 20.8%; Pred. No. 8.2;
 Matches 105; Conservative 53; Mismatches 200; Indels 148; Gaps 22;

QY 20 DKYPLVKWLDSTGTFLAPARNVDVPLDSLOFFIDFKRECKSLGHLPRDLGSPITAFG 79
 Db 348 DKYFVCDQLESPTG-FCLDSAPSHKLSDSSTEFMHIIQVQLGGSTGTSIAP-TPES 405
 QY 80 KICTT---SRRLRLPGEEVEVVGINCRRWLLCAVEKCKWCVHARTLHSGSSL-WE 135
 Db 406 SILSTRKDGRRNSLP-----VKK---TVHPFADLHKDASCND 441
 QY 136 ILYQHSVRLEKRRRRPRFVGENSDSSSEEDHPAFCDVPVQTGAES-----EDSGDEG 189
 Db 442 PFYSSDFRFGSPKATKEL-----SKQDGHVSLAEVSKKCKSEQEARLVLEASDKV 494
 QY 190 PSTRHASGVQPYDDANADSPGSD-----EGPSIHSQDP--PADETTVHTNV 239

Db 495 PE-----STVDGLDAVDAQPSQDSFNGVTGVSQGEKGFPPSSPDGHTLL-ANST 548
 QY 240 EDDLTLDKESACALMYHVGQEMDMWRAMCDLDFLLGIPE-----DVIATSQPG 291
 Db 549 ELKVQLLTVE-----PMWDEYECIPLKASKFNRLVDVASTSQAF 590
 QY 292 GDTDASGVVTEGSIASAAGVAGVEDVVLGALAEQAVAGVEVLE-----ISDEEVD 343
 Db 591 GEPPSSHEKTRGEEGSEHAEKPGKRSKSPRAETEAESRLPEKLPKPPKDPQEDQ 650
 QY 344 GAGLPASRRRPVVGDFLMDGPRRHRPTTRIRHRKLSAYYRVARPPVMTDLRGLVE 403
 Db 651 GHALPP-----ETPADKPKVYEKAKRSKTRHSEEG----- 683
 QY 404 VFVFGPAMSLEVERKVFLCSQN-----PLADISHSCL--HSRKGRLVLLPKPDNNNTG 456
 Db 684 -----EASGFSAVCEEIISAPSTSVSLTSLRSHSEGLD-FKPSFPLSKIS 731
 QY 457 --PDVNL-----LAAVLRSPASG 473
 Db 732 VIPHDLFPYPHYEVPLAAVLEAVAG 757

RESULT 15

LDS_DROME
 ID LDS_DROME STANDARD; PRT; 1061 AA.
 AC P34739; Q9VHY1;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Probable helicase Iodestart.
 GN LDS OR CG2684.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92009170; PubMed=1916263;
 Girdham C.G., Glover D.M.;
 RT "Chromosome tangling and breakage at anaphase result from mutations in Iodestart, a Drosophila gene encoding a putative nucleoside triphosphate-binding protein."
 RL Genes Dev. 5:1786-1799 (1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacheb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Rainet K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
CC -|- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -|- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -|- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 946.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62629; CAA44496.1; ALT FRAME.
DR EMBL; AE003676; AF54167.1; -
DR FRC; A40580; A40580.
DR FlyBase; FBgn0002542; lds.
DR GO; GO:0008094; F:DNA dependent ATPase activity; IDA.
DR GO; GO:0006353; P:transcription termination; IDA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR000330; SNF2_N_
DR Pfam; PF00271; helicase C7.1.
DR Pfam; PF00176; SNF2_N; I.
DR SMART; SMC0487; DEXDC; I.
DR SMART; SMC0490; HELIC; I.
DR PROSITE; PS00690; DEAD ATP HELICASE; FALSE NEG.
KW Hydrolase; Helicase; Nuclear protein; ATP-binding.
FT NP_BIND 465 472
FT SITE 603 606 DEAD BOX.
FT CONFLICT 167 167 R -> G (IN REF. 1).
FT SEQUENCE 1061 AA; 118374 MW; 536BC893B1A90509 CRC64;

Query Match 3.6%; Score 107; DB 1; Length 1061;
Best Local Similarity 22.7%; Pred. No. 8.3;
Matches 82; Conservative 52; Mismatches 149; Indels 78; Gaps 17;

QY 158 ENSDSSEED--HPAFCDVPVTQTG--AESDSGDE-----GPSTRHSAAGVQPv---D 203
Db 82 QSEDESDSDVRSANRKKKPLGIFSDSEDELEQRLSPSTRMSITGVAPQLSD 141
QY 204 DANADSPGSGDEGPFSTRHSDSQPPADE-----TTVHTDNVEDDL-----TLLDK 249
Db 142 DSEIYSDEVOEGF-----TEAFTAENVPRVTTQFAGNIQNDLHSTIGADSEVLDD 195
QY 250 SACALMYHVQEMDMLMRAMCDDELDF-----DLIGIP-EDVIATISQPGDTSAGVTEGS 304
Db 196 SGSDVILSNKETPIELTSDTDATTKENMSPGPPFRPSKLSRPSRAGAVVKTSKN 255
QY 305 IAASAVGAGVEDVYLALGAEQAQNVAGEYVLIS-----DEEV-----DQAGLPPASRRPV 356
Db 256 LSQPTQAVLQKNTSPAAPRRSRIKSDQKVQSVQVYDEEMRKLAEKRVQVSDAEKLF 315
QY 357 VGSEFLWDDGPRRHERPTTR-----TEHRKLRSAYYEVARP-----PWMIT 397
Db 316 VAHKLDPKSGQIKRQIDTLRLRELANDEQWTSALRVQSNVPAVRVVKPTLNPRAVSIDT 375
QY 398 ----DRLG-----VEVFYFGRPAMSLVERXVFILCSQNPILADISHSCLEHKGRLVLPK 449

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 08:57:44 ; Search time 45 seconds
(without alignments)
3975.529 Million cell updates/sec

Title: US-10-656-499-2
Perfect score: 2993
Sequence: 1 MAGRLTWISFIVGALDSD.....QFDVPLVILKRLRSVTCCD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_arches.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2993	100.0	567	12	Q99AM3
2	2977.5	99.5	566	12	Q9DUN1
3	1620	54.1	313	12	O40940
4	828	27.7	152	12	Q9DUM9
5	180.5	6.0	467	12	O40941
6	177.5	5.9	467	12	P88948
7	155.5	5.2	696	12	O40939
8	155.5	5.2	696	12	P88947
9	126.5	4.2	1373	10	Q9AX46
10	125.5	4.2	777	5	O44122
11	125	4.2	777	5	Q9VL91
12	123	4.1	462	13	Q9DFH2
13	122	4.1	364	12	Q9WRN7
14	121.5	4.1	947	5	Q9BIA3
15	121.5	4.1	1033	16	Q8D6Y3
16	121	4.0	361	12	Q9J2U5

17	120	4.0	253	12	Q9WRP1
18	120	4.0	998	12	Q9DND5
19	119.5	4.0	1055	11	Q8CHG2
20	119.5	4.0	1415	11	Q8OTFO
21	118	3.9	163	12	O71294
22	117.5	3.9	626	2	Q8KPN5
23	117.5	3.9	1567	11	O35243
24	117	3.9	1316	4	O75128
25	116.5	3.9	487	13	Q8TFA9
26	116.5	3.9	506	11	Q8OV12
27	116.5	3.9	565	13	Q8JIP0
28	116	3.9	512	10	Q8LTH6
29	115.5	3.9	466	11	Q8C916
30	115.5	3.9	3436	12	Q66666
31	114.5	3.8	672	10	Q851R1
32	114	3.8	691	4	Q9AQF7
33	112.5	3.8	748	10	Q8L734
34	112.5	3.8	748	10	Q9GRT1
35	111	3.7	1161	11	Q8R320
36	110.5	3.7	566	5	Q8T6T5
37	110	3.7	974	5	Q9Y159
38	110	3.7	975	5	Q8GB47
39	109.5	3.7	257	11	Q8CU64
40	109.5	3.7	716	2	Q8RLI5
41	109.5	3.7	853	16	Q8ZUM5
42	109.5	3.7	1246	5	Q9NEE0
43	109	3.6	1026	5	Q9N6D9
44	109	3.6	1664	10	Q9ZOK0
45	109	3.6	4717	3	O94248

ALIGNMENTS

RESULT 1

ID	Q99AM3	PRELIMINARY;	PRT;	567 AA.
AC	Q99AM3			
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)		
DE		B-cell specific latent nuclear protein.		
OS		Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).		
OC		Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
CC		Gammaherpesvirinae; Rhadinovirus.		
OX		NCBI_TaxID=37296;		
RN		[1]		
RP		SEQUENCE FROM N.A.		
RX		MEDLINE=20569372; PubMed=11119611;		
RA		Rivas C., Thlick A.E., Parravicini C., Moore P.S., Chang Y.;		
RT		'Kaposi's sarcoma-associated herpesvirus LAN2 is a B-cell-specific		
RL		latent viral protein that inhibits p53.';		
DR		J. Virol. 75:429-438(2001).		
DR		EMBL; AY008303; AGS0357.1; -.		
DR		GO; GO:0005634; C:nucleus; IEA.		
KW		Nuclear protein.		
SQ		SEQUENCE 567 AA; 62663 MW; 7AF404225638D3DC CRC64;		

Query Match	100.0%;	Score 2993;	DB 12;	Length 567;
Best Local Similarity	100.0%;	Pred. No. 3.6e-228;		
Matches 567;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MAGRRLTWISFIVGALDSDKYPVLVKWLDRTGTFILAPARNNDVILDSLOFFIDFKREC	60	
Db	1	MAGRRLTWISFIVGALDSDKYPVLVKWLDRTGTFILAPARNNDVILDSLOFFIDFKREC	60	
Qy	61	LSKGLHPRDLGSPITAFGKICTTSRRLRLPGEEYEVVQGINCRWRLLCAEVKECWC	120	
Db	61	LSKGLHPRDLGSPITAFGKICTTSRRLRLPGEEYEVVQGINCRWRLLCAEVKECWC	120	
Qy	121	VHARTHLHSGSSWEILYQHSVELEKRRPRPFVGENSDSSEEDHPAFCDVPTQTGA	180	
Db	121	VHARTHLHSGSSWEILYQHSVELEKRRPRPFVGENSDSSEEDHPAFCDVPTQTGA	180	

181 ESDSGDEGSTRHSGVQPVDDANADSPGSDGSTRHSDSQPPADTTVHTDNE 240
 181 ESDSGDEGSTRHSGVQPVDDANADSPGSDGSTRHSDSQPPADTTVHTDNE 240
 241 DDLTLKESACALMYHVGQEMDMMLRAMCDEDLFDLLGIPEDVIATSPGSDTASGV 300
 241 DDLTLKESACALMYHVGQEMDMMLRAMCDEDLFDLLGIPEDVIATSPGSDTASGV 300
 301 TEGSIAASVAGVEDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPASRRRPVVG 360
 301 TEGSIAASVAGVEDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPASRRRPVVG 360
 361 LWDGPRRHERPTTRIRHRKLSAYYVARPPVMTDRLGVEVYFGRPAMSLVERKV 420
 361 LWDGPRRHERPTTRIRHRKLSAYYVARPPVMTDRLGVEVYFGRPAMSLVERKV 420
 421 FILCSQNPDLADISHSCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLR 480
 421 FILCSQNPDLADISHSCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLR 480
 481 SGIYVKNLCKSVLYHGNNPPKFGVIGLSSRAVLDFNVNAQYRIQGHKIKTTVF 540
 481 SGIYVKNLCKSVLYHGNNPPKFGVIGLSSRAVLDFNVNAQYRIQGHKIKTTVF 540
 541 GDPTSABQFDMVPLVILKRLRSVTCDD 567
 541 GDPTSABQFDMVPLVILKRLRSVTCDD 567

RESULT 2
 Q9DUN1 PRELIMINARY; PRT; 566 AA.
 AC Q9DUN1
 DT 01-JAN-2001 (TrEMBLrel. 16, Created)
 DT 01-JAN-2001 (TrEMBLrel. 16, Last sequence update)
 DE Interferon regulatory factor 3.
 GN VIRF-3.
 OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=37296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20392178; PubMed=10933732;
 RA Lubyova B., Pitha P.M.;
 RT "Characterization of a novel human herpesvirus 8-encoded protein,
 RT VIRP-3, that shows homology to viral and cellular interferon
 RT regulatory factors." (2000).
 RL J. Virol. 74:8194-8201 (2000).
 DR EMBL; AF157602; AAF23608.2; -;
 SQ SEQUENCE 566 AA; 62507 MW; 822C557407A8C84B CRC64;

Query Match 99.5%; Score 2977.5; DB 12; Length 566;
 Best Local Similarity 99.8%; Pred. No. 6.1e-227;
 Matches 566; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1 MAGRLTWISFIVGALSDSKYPLVKWLDRTSTGTFLAARNNDVPLDSLOFFIDFKREC 60
 1 MAGRLTWISFIVGALSDSKYPLVKWLDRTSTGTFLAARNNDVPLDSLOFFIDFKREC 60
 61 LSKGLHPRDLGSPITAFGKICTTERRLRLLPGEYEVVQGINCRWELLCAEVKECWC 120
 61 LSKGLHPRDLGSPITAFGKICTTERRLRLLPGEYEVVQGINCRWELLCAEVKECWC 120
 121 VHARTHLHSGSISWELLYQHSVRLKRRRRPRPFVGENSDSSSEEDHFAFCVPTQTGA 180
 121 VHARTHLHSGSISWELLYQHSVRLKRRRRPRPFVGENSDSSSEEDHFAFCVPTQTGA 180
 181 ESDSGDEGSTRHSGVQPVDDANADSPGSDGSTRHSDSQPPADTTVHTDNE 240
 180 ESDSGDEGSTRHSGVQPVDDANADSPGSDGSTRHSDSQPPADTTVHTDNE 239

241 DDLTLKESACALMYHVGQEMDMMLRAMCDEDLFDLLGIPEDVIATSPGSDTASGV 300
 240 DDLTLKESACALMYHVGQEMDMMLRAMCDEDLFDLLGIPEDVIATSPGSDTASGV 299
 301 TEGSIAASVAGVEDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPASRRRPVVG 360
 300 TEGSIAASVAGVEDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPASRRRPVVG 359
 361 LWDGPRRHERPTTRIRHRKLSAYYVARPPVMTDRLGVEVYFGRPAMSLVERKV 420
 360 LWDGPRRHERPTTRIRHRKLSAYYVARPPVMTDRLGVEVYFGRPAMSLVERKV 419
 421 FILCSQNPDLADISHSCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLR 480
 420 FILCSQNPDLADISHSCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLR 479
 481 SGIYVKNLCKSVLYHGNNPPKFGVIGLSSRAVLDFNVNAQYRIQGHKIKTTVF 540
 480 SGIYVKNLCKSVLYHGNNPPKFGVIGLSSRAVLDFNVNAQYRIQGHKIKTTVF 539
 541 GDPTSABQFDMVPLVILKRLRSVTCDD 567
 540 GDPTSABQFDMVPLVILKRLRSVTCDD 566

RESULT 3
 O40940 PRELIMINARY; PRT; 313 AA.
 AC O40940
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Similarity to K10.
 OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=37296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97296220; PubMed=9151804;
 RA Neipel F., Albrecht J.C., Fleckenstein B.;
 RT "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
 RT human herpesvirus 8: determinants of its pathogenicity";
 RL J. Virol. 71:4187-4192 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Neipel F., Albrecht J.-C., Ensaeer A., Huang Y.-Q., Li J.J.,
 RA Friedman-Kien A.E., Fleckenstein B.;
 RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma";
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EXBL; U93872; AAB62648.1; -;
 SQ SEQUENCE 313 AA; 34319 MW; 0069EA68F69500BE CRC64;

Query Match 54.1%; Score 1620; DB 12; Length 313;
 Best Local Similarity 100.0%; Pred. No. 7e-120;
 Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

255 MYHVGQEMDMMLRAMCDEDLFDLLGIPEDVIATSPGSDTASGVVTEGSIASAVGAGV 314
 1 MYHVGQEMDMMLRAMCDEDLFDLLGIPEDVIATSPGSDTASGVVTEGSIASAVGAGV 60
 315 EDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPASRRRPVVGELWDGPRRHERPTT 374
 61 EDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPASRRRPVVGELWDGPRRHERPTT 120
 375 RRIRHRKLSAYYVARPPVMTDRLGVEVYFGRPAMSLVERKVFIICSNPLADISH 434
 121 RRIRHRKLSAYYVARPPVMTDRLGVEVYFGRPAMSLVERKVFIICSNPLADISH 180
 435 SCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVKNLCKSVTL 494
 181 SCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVKNLCKSVTL 240

QY 495 YHGNPPKFGVTCGLSSRAVLDFVNAQYRIQGHHEIKKTVFIQGDPTSAEQFDMVPL 554
Db 241 YHGNPPKFGVTCGLSSRAVLDFVNAQYRIQGHHEIKKTVFIQGDPTSAEQFDMVPL 300

QY 555 VIKLRLSVTCDD 567
Db 301 VIKLRLSVTCDD 313

RESULT 4
Q9DUM9 PRELIMINARY; PRT; 152 AA.
AC Q9DUM9;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE ORF K10.6.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
NCBI_TaxID=37296;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=20392178; PubMed=10933732;
RA Lubyova B., Pitha P.M.;
RT "Characterization of a novel human herpesvirus 8-encoded protein,
RT vIRF-3, that shows homology to viral and cellular interferon
RT regulatory factors."
RL J. Virol. 74:8194-8201(2000).
DR EMBL; AF254765; AAF98389.1; -
SQ SEQUENCE 152 AA; 17801 MW; 9492017B2BCA7076 CRC64;

Query Match 27.7%; Score 828; DB 12; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.2e-57;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGRRITWISERTVGLDSDKYPLVWLDSTGTFLAPARNDVPLDLSQFFIDFKREC 60
Db 1 MAGRRITWISERTVGLDSDKYPLVWLDSTGTFLAPARNDVPLDLSQFFIDFKREC 60

QY 61 LSKGLHPRDLGSPITAFGKICTTSRRRLRPGEEVEVQGCNRRWLLCAEVKECWWC 120
Db 61 LSKGLHPRDLGSPITAFGKICTTSRRRLRPGEEVEVQGCNRRWLLCAEVKECWWC 120

QY 121 VHARTHLHGSSLWEILYOHVSRLEKRRRPR 152
Db 121 VHARTHLHGSSLWEILYOHVSRLEKRRRPR 152

RESULT 5
O40941 PRELIMINARY; PRT; 467 AA.
AC O40941;
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE ORF K11.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
NCBI_TaxID=37296;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97296220; PubMed=9151804;
RA Neipel F., Albrecht J.C., Fleckenstein B.;
RT "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
RT human herpesvirus 8: determinants of its pathogenicity?";
RL J. Virol. 71:4187-4192(1997).
[2]
RN SEQUENCE FROM N.A.
RA Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,
RA Friedman-Kien A.E., Fleckenstein B.;

"The genome of human herpesvirus 8 cloned from Kaposi's sarcoma."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U93872; AB82647.1; -
SQ SEQUENCE 467 AA; 50467 MW; 9B72A0155F325SAC CRC64;

Query Match 6.0%; Score 180.5; DB 12; Length 467;
Best Local Similarity 21.7%; Pred. No. 9.9e-06;
Matches 106; Conservative 72; Mismatches 186; Indels 125; Gaps 22;

QY 167 HPAFCDVP-VTQTGAESDSDGCP-STRSASGVQ-----PVDDANADSPGSGDEGP 217
Db 2 HSLFFEPFSGFGSSQSSLLAPDSFRSTQVQGPLHVHTPD---LCLPTGGGLPSP 58

QY 218 STRHSDSQ---PPPADETTV-----HTDNVEDDLTLDKESACA-LMYHVQDEM 262
Db 59 VIFPHETQGLLAPAGOSQTFPSPEGVPVSHVSGLDCLPMVDHIEGCLLDLSDVQCEL 118

QY 263 -DM--LNRAMCD-----EDLFD 276
Db 119 PDLDGLGELICETASPOGPMQSGEGESTESVLPATHPLESSAPGASVMGSGQLPDP 178

QY 277 LLGIPEDVIATPSQGGTDASGVVTEGS-----IAASAVAGVEDVYLAGALE 324
Db 179 LGDLSLCELCTASQPGPMQSEG-GEESTESVSLPATHPLESSAPGASV---MGSSFQ 233

QY 325 AQNVAGYV-----LEISDEVDGACLPASRRPVVGEFLNDDGRRRHERPTRR--I 377
Db 234 ASDNVDFIDICPELCRDRDVED-----QEKADQTFYWGSDMRPKVLTATOSVA 284

QY 378 RHKRLRSAYYRVARPPVMTDRLGVEYFVFGRPAMSLVERKVFILCSQNLADISHSCL 437
Db 285 AYLKSKQAIYKVGDKLPLV-----VEVYFGEKVKHFDLTGGIVICSOVPEASPEHC- 339

QY 438 HSRKGLRVLLPKP-----DNNTGPGGVNLLAAVLRSFASGLVIVSRGIVYVKNLCKST 492
Db 340 QTVPYKLLLPRTAHCSVDANRT-----LEQTLDRFSMGVVAIGTNMGIKGLLEY 392

QY 493 VLHGNPPKFGVTCGLSSR-AVLDFVNAQYRIQGHHEIKKTVFIQGDPTSAEQFDM 551
Db 393 AYFGNARSRRIGKCRPLSHRHEIQQAFDVERN-REPEGRSYASLFLGRRPSPEYDWDH 451

QY 552 VPLVIKRL 560
Db 452 YFVILHIYL 460

RESULT 6
P88948 PRELIMINARY; PRT; 467 AA.
ID P88948
AC P88948;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE ORF K11.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
NCBI_TaxID=37296;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97094384; PubMed=8939871;
RA Moore P.S., Boshoff C., Weiss R.A., Chang Y.;
RT "Molecular mimicry of human cytokine and cytokine response pathway
RT genes by KSHV."
RL Science 274:1739-1744 (1996).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=97121480; PubMed=8962146;
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
RT (HHV8)."
RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867 (1996).

[3]
SEQUENCE FROM N.A.
RU Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U75698; AAC57142.1;
SQ SEQUENCE 467 AA; 5036 MW; 9E72A7789F5325AC CRC64;

Query Match 5.9%; Score 177.5; DB 12; Length 467;
Best Local Similarity 21.7%; Pred. No. 1.7e-05;
Matches 106; Conservative 71; Mismatches 187; Indels 125; Gaps 22;

QY 167 HPACDVP-VTQTAESEDSGDEGP-STRHSAAGVQ-----PVDDANADSPGSGDEGP 217
Db HSLFEEPEEPGFGSSGSSILLAPDSRPRTSQVQGLVHVTDT---LCLPTGGLPSP 58

QY 218 STRHSDSQ---PPPADETTV-----HTDNVEDDITLLDKESACA--LMTYHVQEM 262
Db VIFPHETQGLLAPAGQSQTFFSPGFPVPSVSGLDCLPMDVHIEGCLLDLLSDVQQL 118

QY 263 -DM--LMRAMCD-----EDLFD 276
Db PDLGDLGELLCTASPOGPMQSEGEGSTESVSLPATHPLESSAPCASVMGSGQLPD 178

QY 277 LLGIPEDVIATSQPGDITDAGVTEGS-----IAAGAVGAGVEDVYLAGALE 324
Db LGDLSSELLCTASQPGPMQSEG-GEESTESVSLPATHPLESSAPCASV-----MGSSFQ 233

QY 325 AQNVAGEVY-----LEISDEVDGAGLPPASRRRPVVGDFLWDGPRRHERPTTR--I 377
Db ASDNVDDIDICPLCRDRDVED-----QEKADQTFYWGSDMRPKVLTATQSA 284

QY 378 RHRKLSAYYVARPPVPMITDRLGVEYFGRPAWSLEVERKVFILCSQNPADISHSL 437
Db AYLKQAIYKVGKLVPLV-----VEVYFGKVKYTHFDLTGGIVICSQVPEASPEHIC- 339

QY 438 HSRKGLRVLPLPKP-----DDNTPGPDVNLAAVLRSPFASGLVIVSLRGIYKVLCKST 492
Db QTPVPYKLLPRTAHCSVDANRT-----LEQTLDRFSMGVAVGNTMGIFLKGLEY 392

QY 493 VLYHGNNPKFVGICGLSSR-AVLVDVNVAAQRIQGHIEKTKTTFVIGDPTSAEQFDM 551
Db AYFCVNASRRRIGKCRPLSHRHEITQAPDVERHN-REPEGRVYASFLGRRPSPDYSDH 451

QY 552 VPLVILKRL 560
Db YPVLHIVL 460

RESULT 7
O40939 PRELIMINARY; PRT; 696 AA.
AC O40939;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ORF K10.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97296220; PubMed=9151804;
RA Neipel F., Albrecht J.C., Fleckenstein B.;
RT "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
human herpesvirus 8: determinants of its pathogenicity?";
RL J. Virol. 71:4187-4192(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,
RA Friedman-Kien A.E., Fleckenstein B.;
RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma.";

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U93872; AAB62660.1;
SQ SEQUENCE 696 AA; 74036 MW; C0B46CCD3934P4A7 CRC64;

Query Match 5.2%; Score 155.5; DB 12; Length 696;
Best Local Similarity 22.9%; Pred. No. 0.0016;
Matches 103; Conservative 58; Mismatches 192; Indels 97; Gaps 22;

QY 151 PRPFPVGENSDSSEEDHPAFCDVPVTQTGAESSEDSGDEGPSTRHSAAGVQVDDANADSP 210
Db PTEPEVAERQEPSSSGIPVVCQ-----GDMNATGYRRVTTSAGALEVE-IIDLTDGS- 352

QY 211 GSGDEGPSTRHSDSQPPPADETTVHTDNVEDDITLLDKESACALMYHVQEMDMLRAMC 270
Db -----DTPSTVA-STFLPVSGPRVFPQTV---LYSAPFAVNPVSHLPTLELER-RECV 403

QY 271 DEEDLFDLIGIPEDVIATSQPGDITDAGVTEGSIAASAVGAGVEDVYLAGALEAQNVA 329
Db PGS-----GERPRVPLVSTYAGDRYAVGVGPEQSLVPPPLGL-----PLTSLN 449

QY 330 GEYVL-----EISDEVDGAGLPPASRRRPVVGDFLWDGPRRHERPTTRIR 378
Db GEDICTWEELGNILSELQEE-----PSSTRQAT-----DRR--RPRSRS 490

QY 379 HRLKLSAYYVARPPVPMI-----TDLRG--VEVYFG--REPMSLEVERKVFILCSQNP 428
Db GRTFVSHSGPEKPSKMFDPDPSQSVFVVEIFVYGNLRTLRREGDAGEAMLCSW-P 549

QY 429 LAD-ISHCLSHRKGL-----RVLLPKPDNNTGPDVNLAAVLRSPFASGLVIVSLRGI 483
Db VGDITGLHLCQSPVPELLRPLRTVPSPE-----QMEILNRFVPEGLGHGPFIFCSMSGI 602

QY 484 YVKNLCKSTVLVHGNNPKFVGICGLSSRAVLVDVNVAAQRIQGHIEKTKT----- 535
Db YSRNATQVEGMVFNPNRKYERILRSFSPVQQLFNTRY-LATTAIPQPLSPVNPVT 661

QY 536 --TVFIGDPTSAEQFDMVPLVILKRLRSV 563
Db CGTVFFGASPASTENFQNVPLVKIFIGSI 691

RESULT 8
P88947 PRELIMINARY; PRT; 696 AA.
AC P88947;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ORF K10.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094384; PubMed=8939871;
RA Moore P.S., Boshoff C., Weiss R.A., Chang Y.;
RT "Molecular mimicry of human cytokine and cytokine response pathway
genes by KSHV";
RL Science 274:1739-1744(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97121480; PubMed=8962146;
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
(HHV8)";
RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

DR	ENBL; UT5698; AAC57141.1; --
SQ	SEQUENCE 696 AA; 74106 MW; 8BEFDCD3934FAA7 CRC64;
	Query Match 5.2%; Score 155.5; DB 12; Length 696;
	Best Local Similarity 22.9%; Pred. No. 0.0016;
	Matches 103; Conservative 58; Mismatches 192; Indels 97; Gaps 22;
QY	151 PRPFVGENSDSSEDDHAFCDVPVTGTGAESDSGDGSPSTRHSASGVQPVDNADNSP 210
DB	302 PTPPEVAERQESSGGIFVQC-----GDNATGYRRVTSSGALEVE-IIDLTGDS- 352
QY	211 GSGDEGPSTRHSDSOPPPADETTVHTDNVEDDLTLIDKESACALMYHVGEWMDLMRAMC 270
DB	353 ----DTPSTTVA-STPLPVSGPRVFQPTV---LYSAPEPAVNPEVSHLFTELER-RECV 403
QY	271 DEDLDLLGIPEDVIATSQPGDFTA-SGVVTGESIAASAVGAGVEDVYLALAEQAQNV 329
DB	404 PGS-----GERPRFLVSTYAGDYAYGVGYGPEOSLVPPPLGL-----PLTSLNLQ 449
QY	330 GWVL-----ETSDSEVDGAGLPASSRRRPFVVGEFLWDGDPHERHPPTTRIR 378
DB	450 GEDITWEEGIGNILSELQES-----PSSSTQQT-----DR--RPASRSPH 490
QY	379 HRKLRSAYRYVARPEVMT-----TDRLG--VEVYFG--RFAMSLEVERKFVILCSNP 428
DB	491 GRRTPVSHSGEKPSPKMFFPDPSQRVSFWELFYVGNLRGTLREGDAGEAMLCSW-P 549
QY	429 LAD-ISHCLSRKGL-----RVLLPKPDNNTPGPDVNLAAVLSRFASGLVIVLSRGI 483
DB	550 VGDTLGHLCQSFPPELLRIPLTVPSP-----QMEILNRVFEGLGHGFIFICMSGI 602
QY	484 YVKNLCKSTVLHYGNPNPKKGTVCGLSRAVLDFVNAQYRTOGHEHIKKT----- 535
DB	603 YSRNATQEVGWFGNPSRYERILRSFSRPVPOQLFNARY-LATTAAIPQPLSVNPT 661
QY	536 --TVFIGDGPTSABQFDMWPLVKILRSV 563
DB	662 CGTVFPGASPASTENFQNVLTKVIFGSI 691

RESULT 9

Q9AX46 PRELIMINARY; PRT; 1373 AA.

ID	Q9AX46	PRELIMINARY; PRT; 1373 AA.
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Putative gag/pol polyprotein.	
GN	P0416D03.20.	
OC	Oryza sativa (Rice).	
CC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OX	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
OX	Ehrhartioideae; Oryzoae; Oryza.	
NCBI_TaxID=4530;		
RN	[1]	

SEQUENCE FROM N.A.
 STRAIN=cv. Nipponbare;
 Sasaki T., Matsumoto T., Yamamoto K.;
 "Oryza sativa nipponbare (GMA3) genomic DNA, chromosome 1, PAC
 clones:P0416D03.";
 Submitted (OCJ-2000) to the ENBL/GenBank/DBJ databases.
 ENBL; AF002872; BAB1132.1; --
 Gramene; Q9AX46; --
 GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0004812; F:tRNA ligase activity; IEA.
 DR GO; GO:006418; P:amino acid activation; IEA.
 DR GO; GO:006310; P:DNA recombination; IEA.
 DR InterPro; IPR001564; Rve.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00665; rve; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 KW Polyprotein.

<p> SQ SEQUENCE 1373 AA; 150143 MW; 7C1A301D051A9B31 CRC64;</p> <p> Query Match 4.2%; Score 126.5; DB 10; Length 1373;</p> <p> Best Local Similarity 23.1%; Pred. No. 0.82;</p> <p> Matches 99; Conservative 148; Mismatches 125; Gaps 23;</p>	<p> QY 144 LEKRRPRPPFVG-----ENSDSS--BEDHP-----AFCDVPTQT 178</p> <p> D5 67 ISKRNRKRYFFGNGIEYDNSGFRNSSESTPRESRPRRLVALPLSLFAAAATDDD 126</p> <p> QY 179 GAREDSGDEG--PSTR-----HSASG--VQPVDDANADSPGSGDEGPSTR-- 220</p> <p> D5 127 AVDADDADDAGHPHRRRRPQHHTADAGHSHRTDDAADADDAG-HWPHRRCRRRLAQA 185</p> <p> QY 221 HSDSQPPAPDE---TVVHTNVDEDDLTLDPKESACALMYHVGOMDMLKAMCDEDFDL 277</p> <p> D5 186 AAASSPVAADKYSTSLATRDITETWTTDDHRSE---RYRDHOKGDTLMAAN-----NL 236</p> <p> QY 278 L-----GIPEDVIATSPGGDTDAGVVTEGSIAA-----SAVGAGVEDVYL 311</p> <p> D5 237 LLVGAAAPRAEACDGAESKLAAL--GTLGLAVSTATALAAAFPAVATGGAANTYY 294</p> <p> QY 320 AGALEAQNAGEVYLE-----ISDEEVDGAGLPPASRRRVPVGEFLWDDGP-----RRHER 371</p> <p> D5 295 NLALAVTLGAVTLIGALVWVSK-----PAARRAAGKLLYAAPPLVATRHRR 345</p> <p> QY 372 PT-----TRIRHR-----KLSAYRYVARPPVMTIDRLGVFVFYGR 409</p> <p> D5 346 RSIQGGAALEDTRNVITARNCSGDLVFPVPATSTHALLAAPTSLMHRRLG-----HLGR 401</p> <p> QY 410 PAMSLYERKVFILCSQNPADISHSCLSHRKGLRVLLPKPDNDNTGFGVNLAAVLRS 469</p> <p> D5 402 EALS-KLIRSSVISCTKDDLPHLCHAC--QLGHHTLPLFSSSSSRASNPFDLIHCDL-- 455</p> <p> QY 470 PASGLVIVS 478</p> <p> D5 456 WTSPIVSVS 464</p>
<p> RESULT 10</p> <p> ID 044122 PRELIMINARY; PRT; 777 AA.</p> <p> AC 044122;</p> <p> DT 01-JUN-1998 (TrEMBLrel. 06, Created)</p> <p> DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)</p> <p> DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)</p> <p> DE Zinc finger 30C.</p> <p> GN ZF30C OR CG3998.</p> <p> OS Drosophila melanogaster (Fruit fly).</p> <p> OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;</p> <p> OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;</p> <p> OC Ephydroidea; Drosophilidae; Drosophila.</p> <p> EN NCBI_TaxID=7227;</p> <p> RX [1]</p> <p> RP SEQUENCE FROM N.A.</p> <p> RC TISSUE=EYE/ANTENNAL DISC;</p> <p> RA Cellforth T., Gaul U.;</p> <p> RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.</p> <p> RL EMBL; AF035275; AAB87630.1;</p> <p> DR FlyBase; FBgn0022720; zf30C.</p> <p> DR InterPro; IPR007087; znf_C2H2.</p> <p> DR Pfam; PF00036; zf-C2H2; 13.</p> <p> DR SMART; SMO0355; znf_C2H2; 12.</p> <p> DR PROSITE; PS00008; ZINC FINGER C2H2_1; 10.</p> <p> DR PROSITE; PSS0157; ZINC FINGER C2H2_2; 11.</p> <p> KW Metal-binding; Zinc; Zinc-finger.</p> <p> SQ SEQUENCE 777 AA; 88280 MW; D1C69524E9E73CA4 CRC64;</p>	<p> Query Match 4.2%; Score 125.5; DB 5; Length 777;</p> <p> Best Local Similarity 22.6%; Pred. No. 0.45;</p> <p> Matches 88; Conservative 108; Mismatches 155; Gaps 23;</p>
<p> OV 58 RECLSKGLHPDL-----LGSPIAFGKICITSRLRLRPLPEEYEVWGIN 103</p>	

```
DB 62 RHCDARFCHBELTQHADEHGVTAQVAGQERNAF--VCEKC-----GAETKYQAYR 112
QY 104 CRRWELLCAEVK-----ECWVCVHARTHLHSGSLWELLYOHSVRLEKRR----- 149
DB 113 -RHCTKGEKLPRESRPMCKCY--TRFSSAN-----LSKRRSRPDT 157
QY 150 --RPARPVG-----ENSDSEEDHPAFCDVPVVTQTGAESDSDGEGPST 192
DB 158 CQPEYDPSGSDGMMKKHAFKPKDRNRSDDED-----TTSEESDSDDIPLA 207
QY 193 RHASGVQPDADANADSPGSDGEGSTRHSQ-----PPADETIVHTDNVEDD 242
DB 208 SRLTKLKQ--ESQNSD---SGDECFDPENNSDDADASGQLPPPA--MVKVEAFDEE 260
QY 243 LTLDKESACALMYHVGEMDMLRAMCDEDLFLL-----GI-PEDVIATQSGGDTA 296
DB 261 ----DFEYQDASMYVKTSTDFISNE--KOKLLDVLNNEGDLKPFPSLVEQ----- 307
QY 297 SGVVTGSIASAVGAGVEDVYLAGALEAQNVAGEYVLEISDEEDVDGAGLPASRRPV 356
DB 308 -----GAGILDE--IAAVPLVEVAEDVLELRGHQMEK-----PP----- 340
QY 357 VGEFLWDDGPRRHRPTTRRI-----RHRK 381
DB 341 -----GPKRGPRPKKIPVVKRYK 362
RESULT 11
Q9VL91 ID Q9VL91 PRELIMINARY; PRT; 777 AA.
AC Q9VL91;
CT 01-MAR-2000 (TREMELrel. 13, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Zf30C protein (Lb23102p).
GN ZF30C OR C63998.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RX MEDLINE=2019606; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe C.M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pacibus B., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy L., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
```

```
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J., Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN [2]
SCIENCE 287:2185-2195(2000).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF52805.2; -
DR EMBL; AY051681; AAK93105.1; -
DR FlyBase; FBgn002720; zf30C.
DR InterPro; IPR007087; znf C2H2.
DR Pfam; PF00096; zf-C2H2; 13.
DR SMART; SM00355; ZNF C2H2; 12.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 777 AA; 88347 MW; ED12A39DA882B994 CRC64;
Query Match 4.2%; Score 125; DB 5; Length 777;
Best Local Similarity 23.5%; Pred NO. 0.49; Gaps 20;
Matches 80; Conservative 35; Mismatches 94; Indels 132;
QY 93 GBEYEVVGINCRRWLLCAEVK-----ECWVCVHARTHLHSGSLWELLYOHSVR 143
DB 102 GAETKYQAYR-RHCTKGEKLPRESRPMCKCY--TRFSSAN----- 146
QY 144 LSKHRR-----RPRRPVG-----ENSDSEEDHPAFCDVPVVTQTGA 181
DB 147 LSKHRRRPDTGQPEYDPSGSDGMMKKHAFKPKDRNRSDDED-----TTSEE 196
QY 182 SEDSGDEGSTRHSASGVQPDADANADSPGSDGEGSTRHSQ-----PPADE 231
DB 197 SEDSDDDIPLASRLTKLKQ--ESQNSD---SGDECFDPENNSDDADASGQLPPPA-- 250
QY 232 TTVHTDNVEDDLTLDKESACALMYHVGEMDMLRAMCDEDLFLL-----GI-PEDVI 285
DB 251 -MVKVEAFDEE----DFEYQDASMYVKTSTDFISNE--KOKLLDVLNNEGDLKPFPSL 303
QY 286 ATSPGSGDTASGVVTGSIASAVGAGVEDVYLAGALEAQNVAGEYVLEISDEEDVDGA 345
DB 304 KVEQ-----GAGILDE--IAAVPLVEVAEDVLELRGHQMEK-- 338
QY 346 GLPPASRRRPVVGEEFLWDDGPRRHRPTTRRI-----RHRK 381
DB 339 --PP-----GPKRGPRPKKIPVVKRYK 362
RESULT 12
Q9DFH2 ID Q9DFH2 PRELIMINARY; PRT; 462 AA.
AC Q9DFH2;
CT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE HMG box transcription factor Sox9a.
GN SOX9A.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
```

OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chiang E.F., Pai C.-I., Van Y.-L., Postlethwait J., Chung B.-C.;
RT "Two Sox9 Genes on Duplicated Zebrafish Chromosomes: Overlapping
RT Expression Sites in Embryos but Distinct Expression in Adults";
RL Submitted (Jun-2000) to the EMBL/GenBank/DBJ Databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ab, and tue;
RX PubMed=12397114;
RA Van Y.-L., Miller C.T., Nissen R., Singer A., Liu D., Kira A.,
RA Draper B., Willoughby J., Morcos P.A., Amsterdam A., Chung B.C.,
RA Westerfield M., Hafler P., Hopkins N., Kimmel C., Postlethwait J.H.;
RT "A zebrafish sox9 gene required for cartilage morphogenesis";
RL Development 129:5065-5079(2002).
DR EMBL: AF277096; AAG09814.1; -
DR EMBL: AY090034; AAM13695.1; -
DR EMBL: AY090035; AAM13696.1; -
DR HSP: Q05066; IHR1.
DR ZFIN: ZDB-GENE-001103-1; sox9a.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000910; HMG_12_box.
DR Pfam: PF00505; HMG box; 1.
DR SMART: SM00398; HMG; 1.
DR PROSITE: PS01118; HMG_BOX_2; 1.
SQ SEQUENCE 462 AA; 51333 MW; 78D10994FB67FAFC CRC64;

Query Match 4.1%; Score 123; DB 13; Length 462;
Best Local Similarity 21.9%; Pred. No. 0.34;
Matches 93; Conservative 40; Mismatches 134; Indels 158; Gaps 23;

QY 47 LSLQSFIFKRECLSKGLHP---ROLLSPITAFKICTTTRRLRLPGBE----- 95
DB 4 LDPYKMTDEQKCLSDAFSPMSSEDSAGSPSPSAGSDTENTR---PAENSLLAADGT 59
QY 96 -----YEVVQGINCRWELLCAYKE-----CW 118
DB 60 LGDFKDEEDKFPVCTREAVSQVKGVD---WTLVPMFVRVNGSSKNKPHVPRMFAFW 116
QY 119 WCYHAR-----THLHSG-----SLWEILYQHSVRLEKRRRPRPFVGENSD---S 162
DB 117 WAQARARLADQYPLHNAELSKTLGLMRLNE---VEK-----RPFVEAERLRVQ 166
QY 163 SEEDHAFCDVP-----VTCGAESEDSGDE---GPSTRHSAGVQVDDANADSPGS-- 212
DB 167 HKKHDPDYKYQPRRRKSVKNGSESEDSGSETHISPNALFKAL-----QQADSPASM 219
QY 213 GD-EGPSTRHSQPPPADETTVHTDNVEDDLTLDKESACALMYHVGQEMDMLRAMCD 271
DB 220 GEVHSFSEHSQSQGPTPTPTTKTQPKADL--KREARPLQENTGRPLSINFQ---D 274
QY 272 EDLFDLIGIPEDVIATSQ-----PGDPTDASGVVTEGSTAASAVGAGVEDVYLAGA 322
DB 275 VDIGEL---SSDVETFDVNEFDQYLPFNHQNQ-----PYAGG---YAAWM 315
QY 323 LEQNVAGEVVLSEIDVEDDAGLPPASRRRPPVGEFLWDDGPRRHERPTTTRIRHRLK 382
DB 316 TKFQN-----GSPQSSQLTEJ-----NPAEPDQPRTHIKTEQL 349
QY 383 RSAYY 387
DB 350 SPSHY 354

RESULT 13
Q9WZM7 PRELIMINARY; PRT; 364 AA.
AC Q9WRN7
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE R13 (VIRF).
OS Macaca mulatta rhadinovirus 17577, and
OS Macaca mulatta rhadinovirus 26-95.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gamaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=83534, 119193;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Macaca mulatta rhadinovirus 17577;
RX MEDLINE=99174001; PubMed=10074154;
RA Searles R.P., Berguam E.P., Axthelm M.K., Wong S.W.;
RT "Sequence and genomic analysis of a Rhesus macaque rhadinovirus with
RT similarity to Kaposi's sarcoma-associated herpesvirus/human
RT herpesvirus 8";
RL J. Virol. 73:3040-3053(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Macaca mulatta rhadinovirus 26-95;
RX STRAIN=MACACA MULATTA RHADINOVIRUS ISOLATE 26-95;
RA Alexander L., Denekamp L., Knapp A., Auerbach M.R., Damania B.,
RA Desrosiers R.C.;
RT "The primary sequence of rhesus monkey rhadinovirus isolate 26-95;
RT sequence similarities to Kaposi's sarcoma-associated herpesvirus and
RT rhesus monkey rhadinovirus isolate 17577";
RL J. Virol. 74:3388-3398(2000).
DR EMBL: AF083501; AAD21391.1; -
DR EMBL: AF210726; AAF60043.1; -
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003700; F:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR001346; IRF.
DR Pfam: PF00605; IRF; 1.
DR SMART: SM00348; IRF; 1.
SQ SEQUENCE 364 AA; 41189 MW; D9A48195C39DD018 CRC64;

Query Match 4.1%; Score 122; DB 12; Length 364;
Best Local Similarity 22.5%; Pred. No. 0.29;
Matches 82; Conservative 43; Mismatches 127; Indels 112; Gaps 19;

QY 246 LDKESACALMYHVGQEMDMLRAMCDEDLFDLLGIPEDVIATSPQPGDPTDASGVVTEGSI 305
DB 57 LDRLECGQRQHVSECKRLRLVLRNAGFE---QDDARATTTFRG----- 99
QY 306 AASAVGAGVEDVYLAGALEAQNVAGEVVLSEIDREV-----DQAGLPPASRR 353
DB 100 -----GERFFYLRLPAVDPLCYA--CILDSHSETVLNLEAACVHGLEPGTPLPP--- 146
QY 354 RPVVGELWDDGPRRHERPTTTRIRHKLRSAYVRAR-----PPVMTD--RLGVEV 404
DB 147 -PAPAE---ADG-----AARSVYARAARLATVAPPDQITFFWRLIRV 187
QY 405 FYGGRPAMSLEVE-----RKVFILCSQNPPLADISHSLHSRKLRLVLPKP--DNNTGP 457
DB 188 FYFG-----SLVAETSQRRGRVRLHKQDPKP--GHECFYV-TAYKMWLPKPLDGLPLTP 240
QY 458 GDNVLLAAVLASPASGLVIVSLRSGLIYKVLCKSTVLYHGNP-----PKKF 504
DB 241 EQRETVCIELNGCEGVFLHGNELGMVVDNRTRHTRVACAGNDAEGNHAQRAVRSSVKSQI 300
QY 505 GVICGLSSRAVLVDVNAQYRIQOHE-HIKTTVFIGDPTSAEQFDMVPLVILKRLRSV 563
DB 301 FYVMGLLR-----LARSVPDGTVPFNSAVTLVGLGRPGSKR-PQVPVTL-----V 346
QY 564 TCDD 567
DB 347 ICQD 350

RESULT 14
Q9BIA3 PRELIMINARY; PRT; 947 AA.
ID Q9BIA3
AC Q9BIA3;

01-JUN-2001 (TrEMBLrel. 17, Created)
 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN C25H3.11.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Psolodinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology, the C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX Johnson D., Wilson R.;
 RA "The sequence of C. elegans cosmid C25H3.11";
 RT Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX Waterston R.;
 RA "Direct Submission";
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U95355; AAK31456.1; -;
 DR WormPep; C25H3.11; C827091.
 KW Hypothetical protein.
 SQ SEQUENCE 947 AA; 107680 MW; B037D066D49FD29 CRC64;

Query Match 4.1%; Score 121.5; DB 5; Length 947;
 Best Local Similarity 20.0%; Pred. No. 1.2;
 Matches 107; Conservative 83; Mismatches 193; Indels 152; Gaps 25;

QY 69 DLGSPITAFGKICTTSSRLRLPGEYEVVQGINCRWRLLCA---EVKECWVCVHAR- 124
 Db 283 DILGLLTHR-----QMTOLRAVNEWARFERARQHRKWRPLCTIGENAKWKFAYNRV 337
 QY 125 ---THLSSGSSWEILYQHVSRLKXRRPRPRPVGNSDSEHEDHAPCDVPVQTGAB 181
 Db 338 LBSRRANRRWDFARASGRFNAYCRAYKRLIG-----LIANPNAIQ 382
 QY 182 SEDSGDEGPTSRHSASGVQVDDANADSPGSGDEGPTSRHSQPPADPTTVHTDNVED 241
 Db 383 STVSTDDSNTPSNALAVVPVTSN-----GASSNHQSL-----ESTAIMKQIER 428
 QY 242 D-----LTLLDKSACALMYHVCQEMDMLMRAMCDEDLFDLLGIPEDVIATSPQGGDT 294
 Db 429 DAQYTHELHLPRFVFRKLREKEKELGITAAPIDTDAFETLEPPDDEIIVDEPTAFT 488
 QY 295 DASGVVTEGSIASAVG-----AGVEDVYLALGAQAQNV--AGEYVLE 335
 Db 489 EP-----TNGLYGWITGFFGQAQDEKQEDKDFGNVDV---GELKDINVKEMEIEILD 541
 QY 336 ISDEEVDGAGLPPASRRPVVGEFLMDGPRRHERPTRIRHRKLKLR-----SAYYRVAR 391
 Db 542 VLHESWDDSTLL-----RDALLAQI-----SLRLEHLTRFVDSVDVHDGIEQ 584
 QY 392 PPMITDLRGVEVYFGRPA---MSLEV---ERKVFILCSQNP-----LADISHSLHS- 439
 Db 585 QVLALELGSVSRWELSPKQHYLSVDVTVMDSVQRLSGHPRKSPKPAELSESLLYST 644
 QY 440 RKGLRVLLPKPDNNTPGPDVNLAAVLRSFASGLVIVLSRSIYKNL--CK---STVL 494
 Db 645 ASSTKML-----TVRGEGTDVLSRRKPMFMYIRKSPR--LIVAKHWNCRLRVSV 696
 QY 495 YHGNPPKPKGVLGSLGRAVLVDVFNVAQYRIQGHIEHKTTVFIGDPTSAEQF 549
 Db 697 Y-----EKGALGLS-----TLF-SDPTVFDEF 719

RESULT 15

Q8D6Y3
 ID Q8D6Y3 PRELIMINARY; PRT; 1033 AA.
 AC Q8D6Y3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Conserved hypothetical protein.
 GN VV20388.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.E.;
 RT "Complete genome sequence of Vibrio vulnificus CMCP6";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE016809; AAO07344.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1033 AA; 111627 MW; 7317D4ABE24C35E9 CRC64;

Query Match 4.1%; Score 121.5; DB 16; Length 1033;
 Best Local Similarity 20.1%; Pred. No. 1.4;
 Matches 99; Conservative 69; Mismatches 188; Indels 137; Gaps 22;

QY 13 IVGALSDSKYPLV-----KWLDRSTGTFL-----APAAANDVPLDLSLQFIDFKRE 59
 Db 73 VTGVDPESYPIVVKAIKGTVDGTDGNTIATDYMSAPAGEQDITPLSTVHVLLEDE 132
 QY 60 CLAK-----GLHPRDLGSPV-----TAFG-KICTTSSRLRLPGEYEVVQV 101
 Db 133 TLTKDAVQVATQGLGTSDDVLGYIEDNDVEAAFGAKTLVSSGVLPEETPEE----- 185
 QY 102 INCRRWLLCAEVCWCVHARTHLHSGSSWEILYQHVSRLKXRRPRPRPVGNS 160
 Db 186 -----LASEADE-----ETTTSTFTTEATQVNTYKHEIETEKALGE 226
 QY 161 DSSEEDHPAFCDVPVVTQTGAESDSESGDEGPTSR-----HSASGVQVDDANADSPGSGD 214
 Db 227 ELNLDKVGTFD--PVTGTVTFEEDSDGCVANSQDAPDNSEEMLDSGDDIGDADTDD 285
 QY 215 EGPSTRH-SDSQPPADETT-----VHTDNVEDLTLDKSACALMYHVCQEMDML 265
 Db 286 DNDGTLVDVDAFPDABETITDDGIGNVHTDNDGTLDDTDDAPPLN----- 335
 QY 266 MRAMCDEDLFDLLGIPEDVIATSPQ-----GUTDASGVVTEGSIASAVGAGVEDVILAG 321
 Db 336 -----PEETVDTDKGVGNADTDD--NDGALDG-----DDAPPLN 370
 QY 322 ALBAQNVAGYVLEISDEEVDGAGLPPASRRPVVGEFLMDGPRRHERPTRIRHRK 381
 Db 371 PEETDITDKGIGNADTD-DNDGILVDVDSNPTVPDL-----NIEQVIQFMQ 419
 QY 382 LRSAVYRVARPPVMTDRLGVEVYFGRPAMSLVERKVFILCSQNPDLADISHSLHSR 441
 Db 420 NNSMFYALMADH-EYNDATGTE-----SVETIYVEKFTLA--NNIGVTVEAYQMLPD 467
 QY 442 GLRVLLPKPDNN 454
 Db 468 G-RKVADEPDAND 479

Search completed: June 9, 2004, 09:02:37
 Job time : 48 secs